DEEP LEARNING FOR KOREAN NLP

정상근
SKT
2015-09-22

Outline

- Natural Language Processing
- Deep learning for NLP
  - Symbols to Vector
  - POS Tagging
  - Domain Entity Extraction
  - Intention Analysis
- Recent movements
  - Encoding- Decoding Approach
  - Attention Modeling Approach

Overview

NATURAL LANGUAGE PROCESSING

목적

무엇을 위해 자연어 처리를 할까?

자연어로 만들어진
모든 데이터에 대해서

이해하고

- 문서 이해
- 발화 이해
- 질문 이해
- …

답하기

- 검색
- 추론
- 분류
- …

서울역 근처 스타벅스로 안내해줘

✓ '스타벅스'라는 '장소'
✓ '서울역' '주변' 일 것
✓ '안내'하라는 것

서울역 근처 **스타벅스**로 안내해줘

*Q : 어떻게 **장소**인 단어를 알아낼까?*

- 장소는 보통 '명사' 인 경우가 많음
- ~로, ~까지, ~에 등의 '조사'로 수식 되는 경우가 많음
- …
- …

품사 분석
(Part of Speech Tagging)

- 단어를 기능, 형태, 의미에 따라서 명사, 대명사, 수사, 조사 등등으로 분류

품사 분석
(Part of Speech Tagging)

서울역/NNP 근처/NNG 스타벅스/NNP 로/JKB

안내/NNG 해줘(하/XSV, 아/EC, 주/VX, 어/EF)

N(명사)	NNP	고유명사
	NNG	일반명사
J(조사)	JKB	조사
X(접사)	XSV	동사파생접미사
E(어미)	EC	연결어미
	EF	종결어미
V(용언)	VX	보조용언

세종 코퍼스 기준
http://ithub.korean.go.kr/user/guide/corpus/guide1.do

영역 개체 추출
(Domain Entity Extraction)

서울역/TS 근처/DIST_CLOSE 스타벅스/POI

TS	역
DIST_CLOSE	근접도
POI	Place of Interest

서울역 근처 스타벅스로 안내해줘

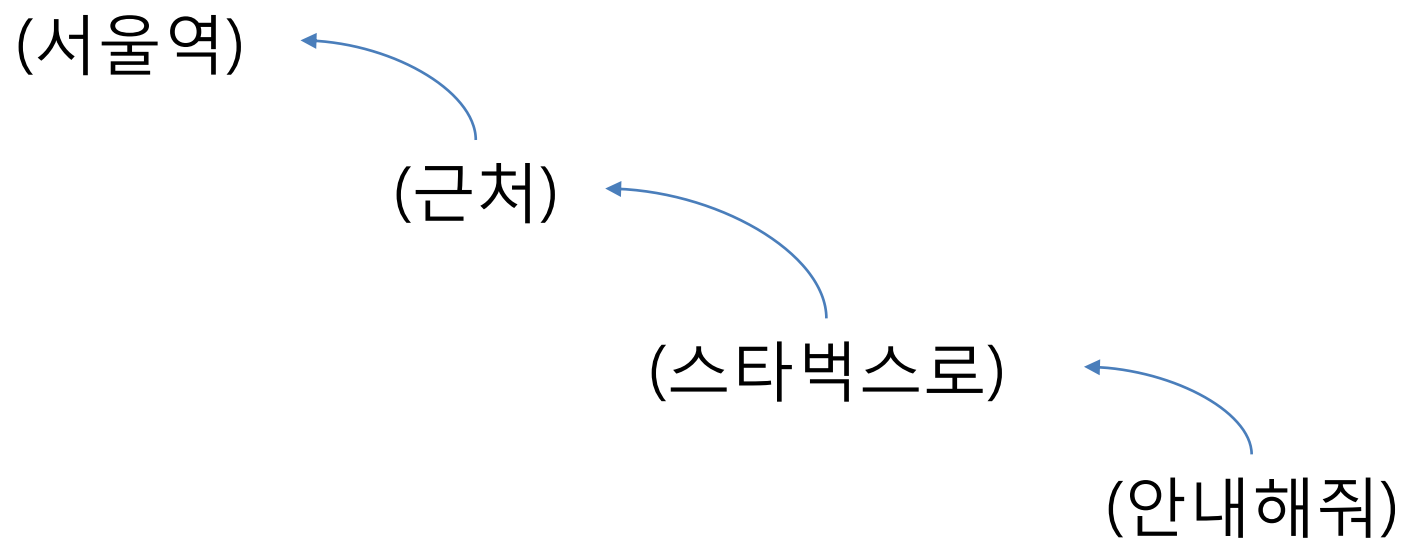
Q : 어떻게 서울역 주변의 스타벅스인것을 알 수 있을까?

- 서울역 ~ 주변 ?
- 주변 ~ 스타벅스 ?
- 이 관계에서 유추

의존 구문 분석
(Dependency Parsing)

- 문장 구성요소들의 관계를 분석

의존 구문 분석 (Dependency Parsing)



- 각 어절(띄어쓰기 기준)이 어디에 '의존' 하는지를 분석
- 품사분석 결과물이 영향을 미침

서울역 근처 스타벅스로 **안내해줘**

*Q : 어떻게 **안내해달라**는 의도를 알아낼까?*

- ~~ 해줘라는 청유형이 '의도'를 담고 있음
- 안내라는 키워드가 존재함
- …

의도 분석
(Intention Analysis)

- 문장의 의도 분석

의도 분석
(Intention Analysis)

서울역 근처 스타벅스로 안내해줘

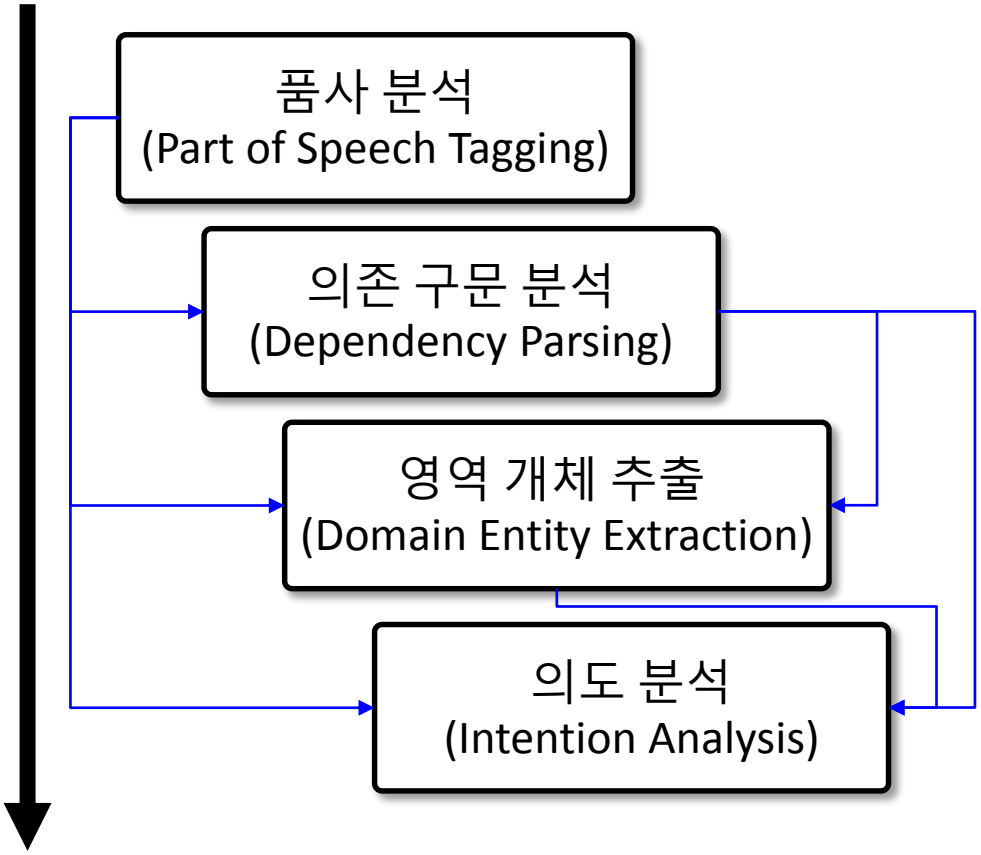
= Set. Destination

- 의도 분석 Tag 는 도메인/서비스/개발자 마다 다름
- 표준화된 분석 Tag가 존재하지 않음

- 단 응용과 상관없는 문장의 역할을 정의하는 기준은 있음
  (Speech Act- https://en.wikipedia.org/wiki/Speech_act)

서울역 근처 스타벅스로 안내해줘

➔ 복수의 자연어 분석 기술이 복합적으로 작용됨



✓ '스타벅스'라는 '장소'
✓ '서울역' '주변' 일 것
✓ '안내'하라는 것

Deep Learning for NLP

SYMBOLS TO VECTOR

Deep Learning 처리 단위 = Vector

Symbol 이 아닌 Vector 가 처리의 기본 단위

[Representation]

Cat

One-Hot Representation

[0, 0, 0, **1**, 0, …]

Cat

Distributed Representation

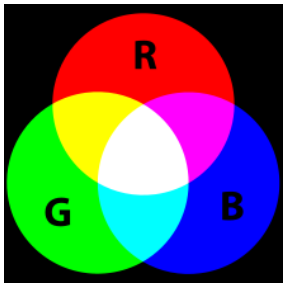
[**34.2, 93.2, 45.3**, …]

비교 – 영상 / 음성 / 자연어

Object		Vector

[영상]

Analog / Digital Converter



RGB = [32,54,11]

[음성]

Analog / Digital Converter

Wave = [12,35,45,10,…]

[자연어]

문자화

"사과"

Vector 가 아님!

자연어의 경우 이미
한번의 '*Digital*'화가
이루어졌다고도 볼 수 있음

자연어의 Vector 화

Object

Vector

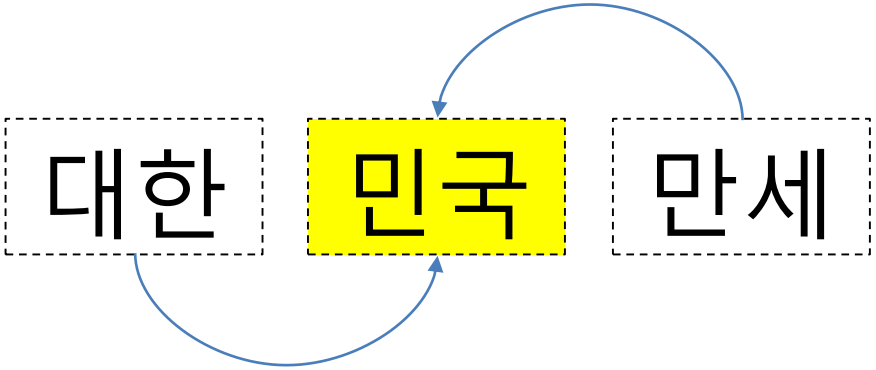
어떻게 자연어를 유의미한 Vector 로 변환 할 것인가?

- Neural Language Modeling
- Word2Vec(Skipgram, CBOW)
- Glove
- Sentence2Vec
- Doc2Vec
- ...

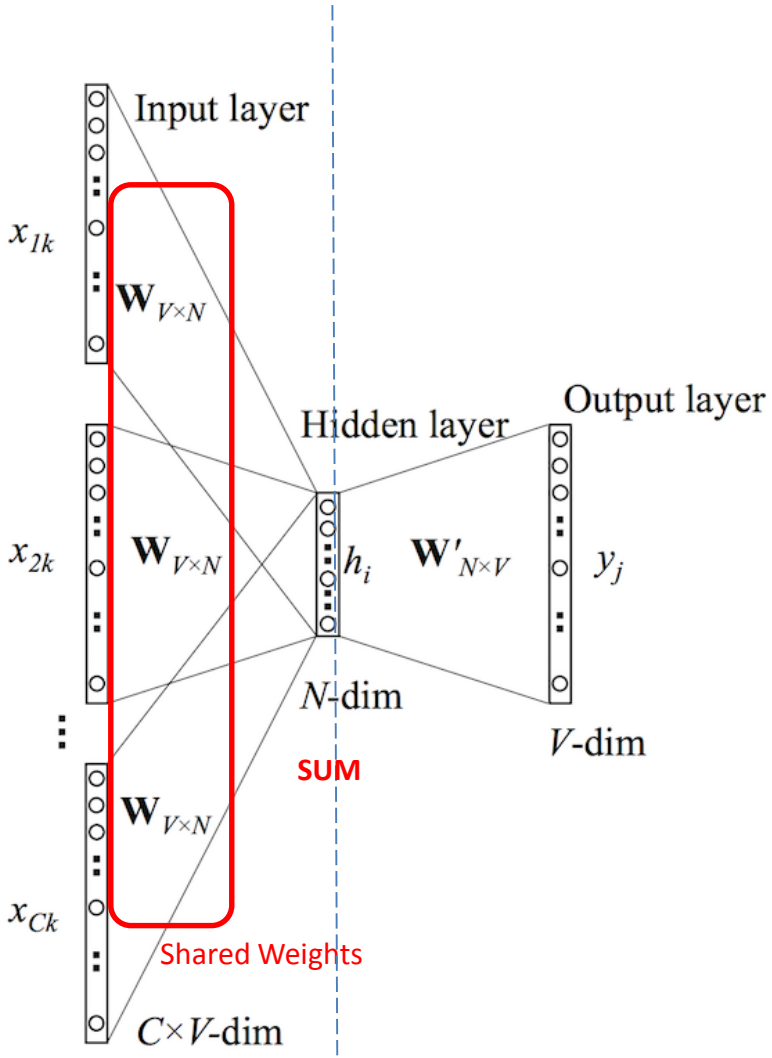
Word2Vec : Cbow

[Idea]

‘단어’란 주변의 단어로 정의된다.



Continuous Bag-of-words Architecture



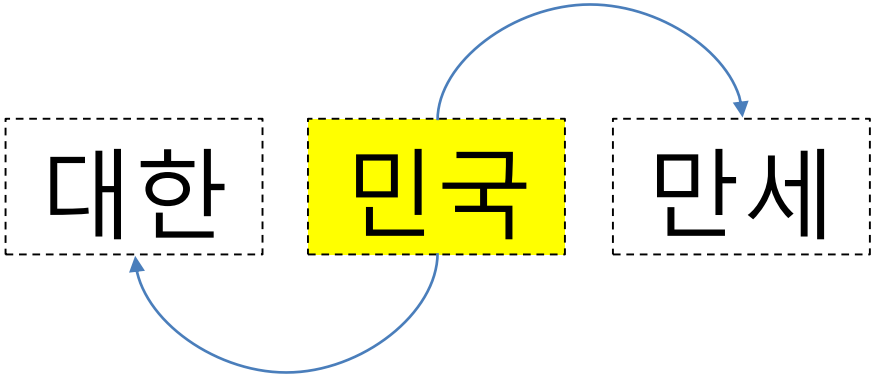
- Predicts the current word given the context

http://alexminnaar.com/word2vec-tutorial-part-ii-the-continuous-bag-of-words-model.html

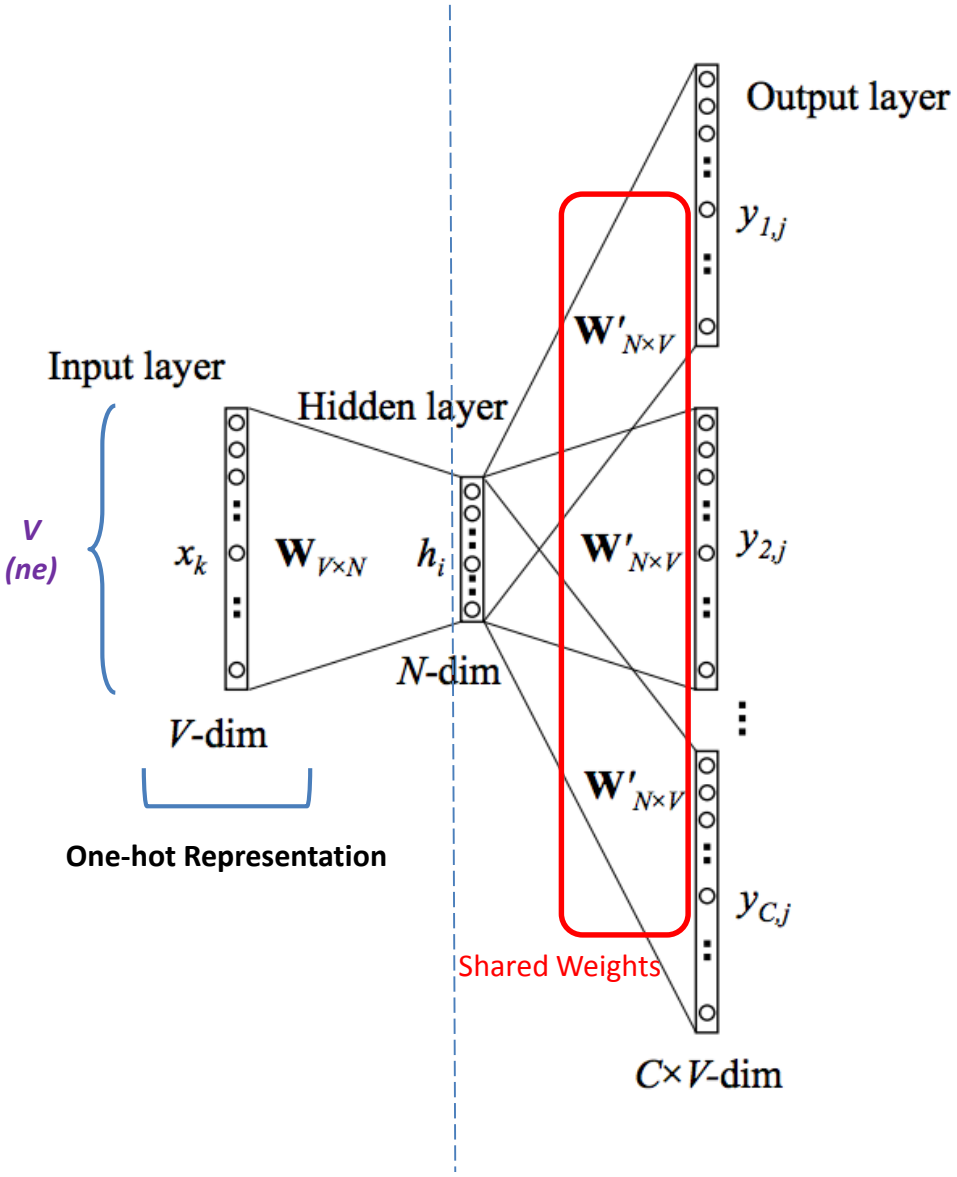
Word2Vec : Skipgram

[Idea]

주변의 '단어'를 잘 설명하는 무엇이
그 단어를 정의한다.



Skip-gram Architecture



- Predicts the surrounding words given the current word

http://alexminnaar.com/word2vec-tutorial-part-ii-the-continuous-bag-of-words-model.html

Word2Vec 결과물 : Semantic Guessing

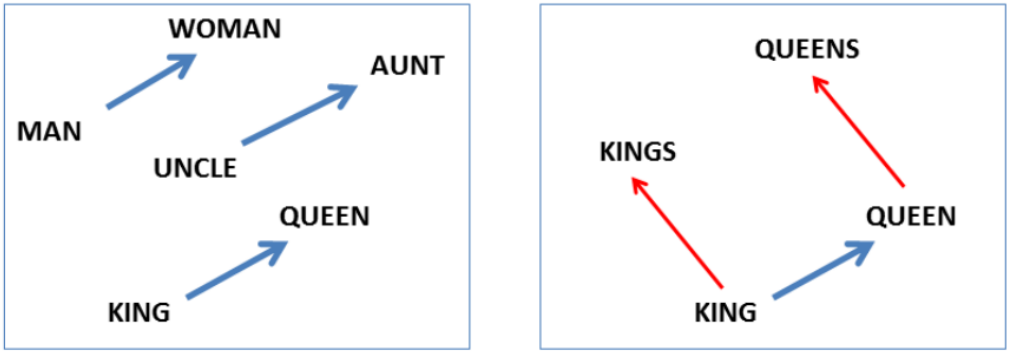


Figure 2: Left panel shows vector offsets for three word pairs illustrating the gender relation. Right panel shows a different projection, and the singular/plural relation for two words. In high-dimensional space, multiple relations can be embedded for a single word.

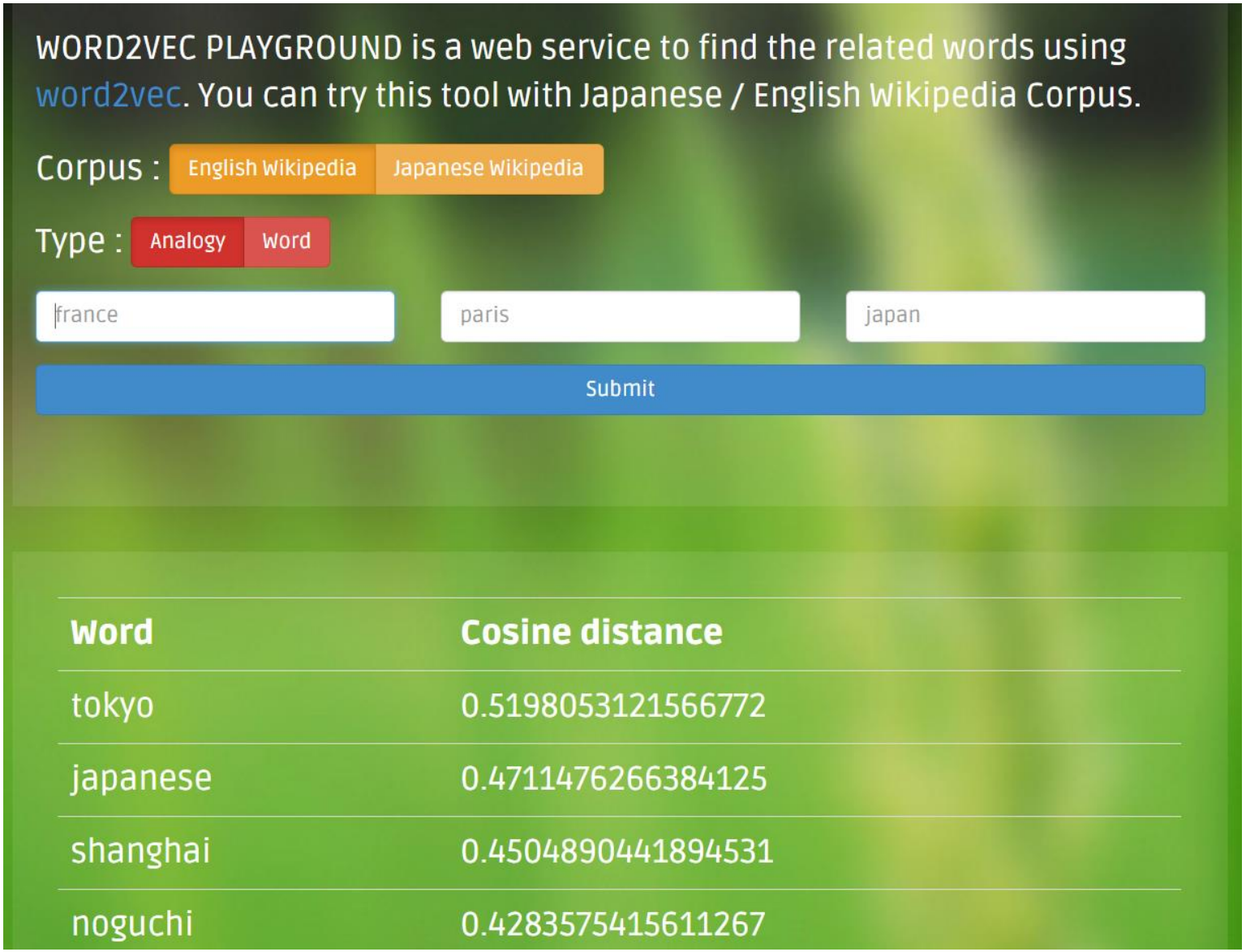
:: DNN 을 통해 Symbol 을 공간상에 Mapping 가능하게 됨으로써 Symbol 들 간의 관계를 '수학적' 으로 추측해 볼 수 있는 여지가 있음

Ex) King – Man + Woman ≈ Queen

:: List of Number 가 Semantic Meaning 을 포함하고 있음을 의미

Microsoft, "Linguistic Regularities in Continuous Space Word Representations", 2013

Word2Vec Demo : Semantic Guessing



http://deeplearner.fz-qqq.net/

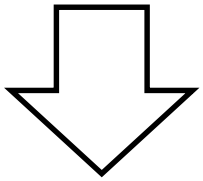
Ex) korea – kimchi
china - ?

Deep Learning for NLP

PART OF SPEECH TAGGING
DOMAIN ENTITY TAGGING
INTENTION ANALYSIS

Sequential Tagging - POS

서울역	근처	스타	벅스	로	안내	해줘
NNP	NNG	NNP		JKB	NNG	XSV~EF



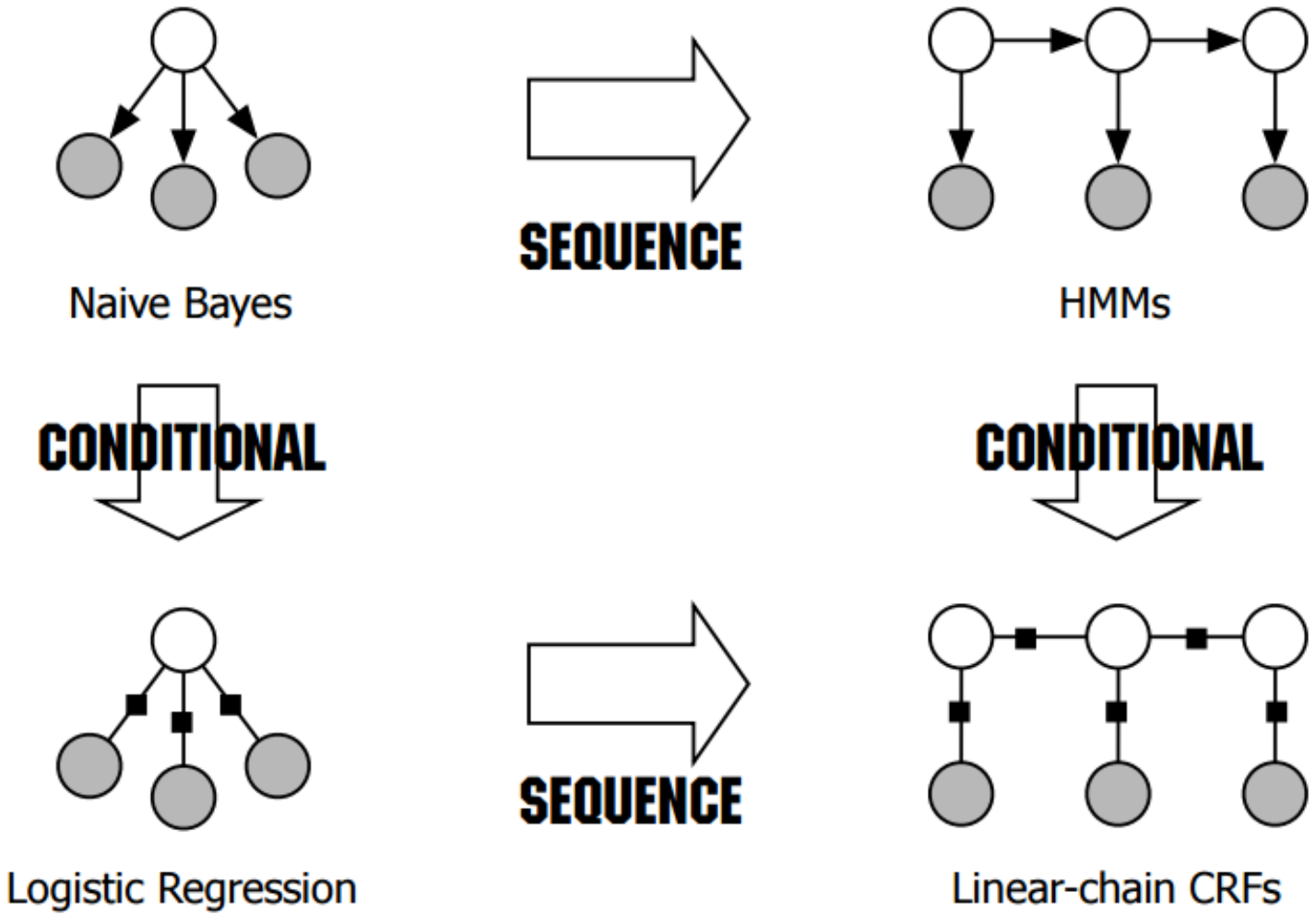
BIO Tagging
(B/Begin, I/Intermediate, O/None)

서울역	근처	스타	벅스	로	안내	해줘
B-NNP	B-NNG	B-NNP	I-NNP	B-JKB	B-NNG	B-XSV~EF

Sequential Tagging – Domain Entity

서울역	근처	스타	벅스	로	안내	해줘
B-TS	B-DIST_CLOSE	B-POI	I-POI	O	O	O

(Traditional) Sequential Tagging Method



"An Introduction to Conditional Random Fields for Relational Learning", Charles Sutton and Andrew McCallum

Sequence Modeling

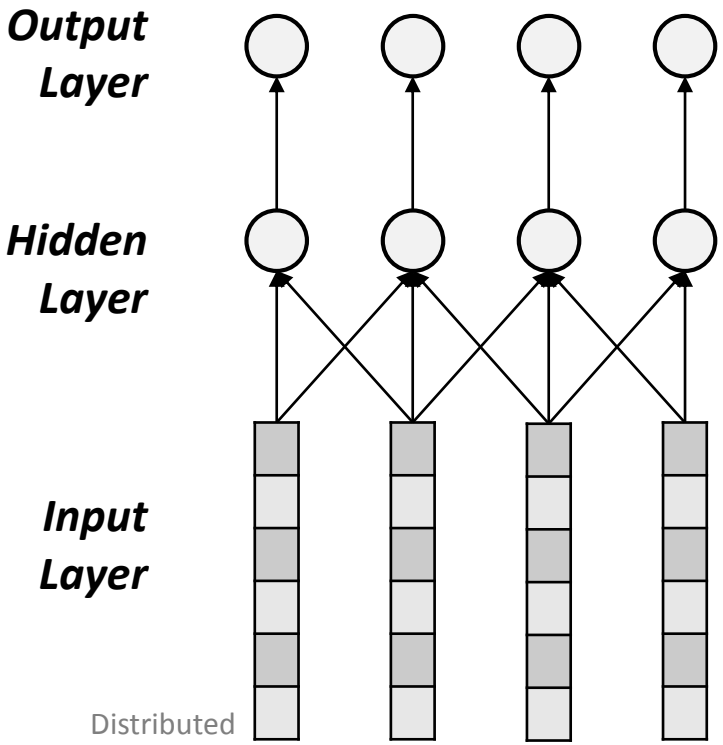
@ Deep Learning

Sequence 를 어떻게 다룰 것 인가?

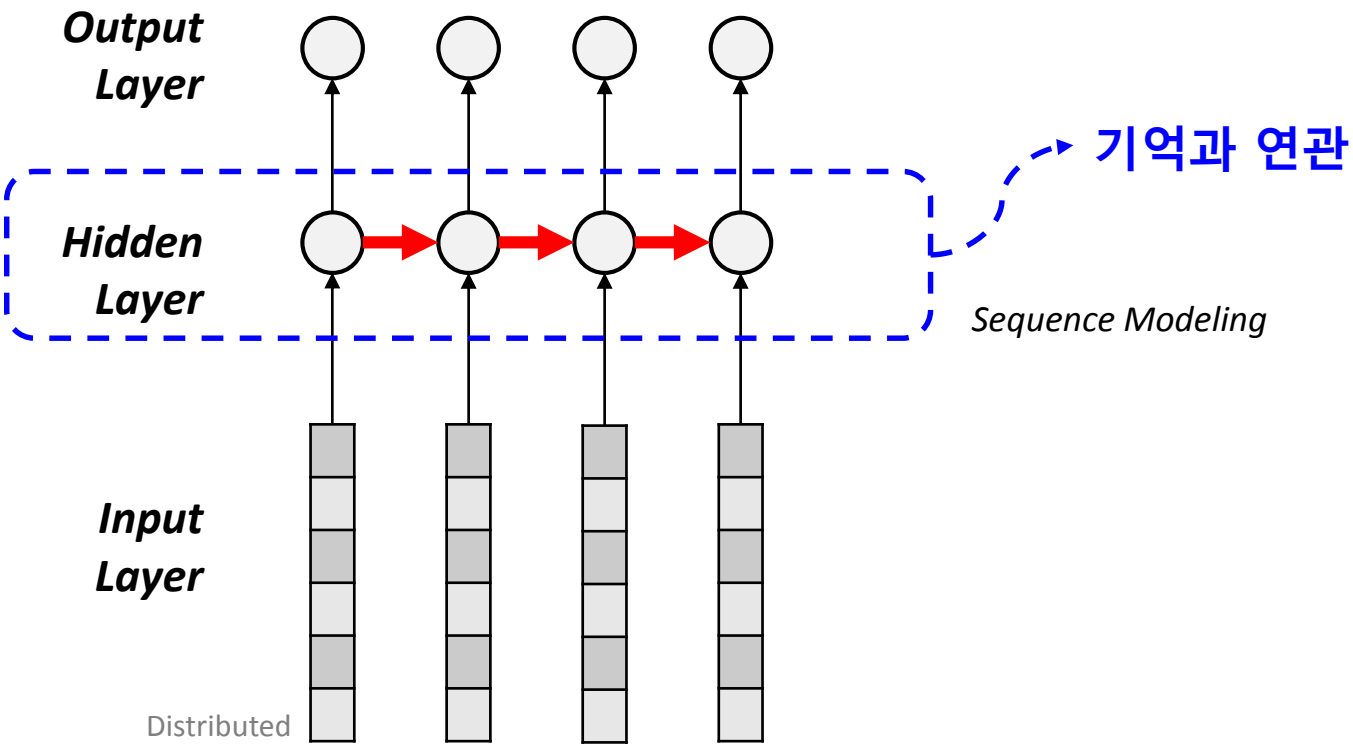
≈

여러 번 수행	Sequence를 모델링

여러 번 (독립적으로) Tagging

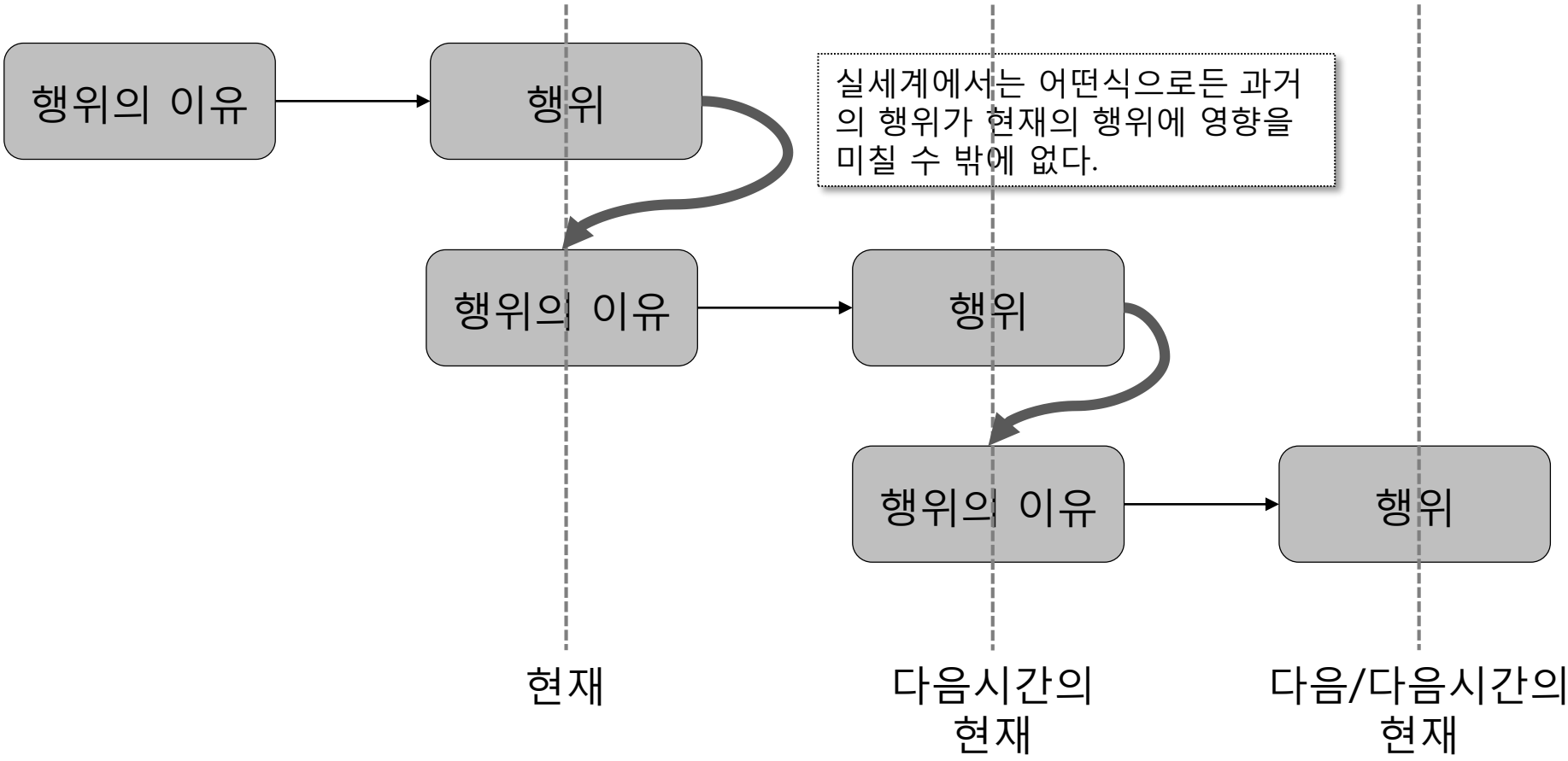


Sequence 를 모델링

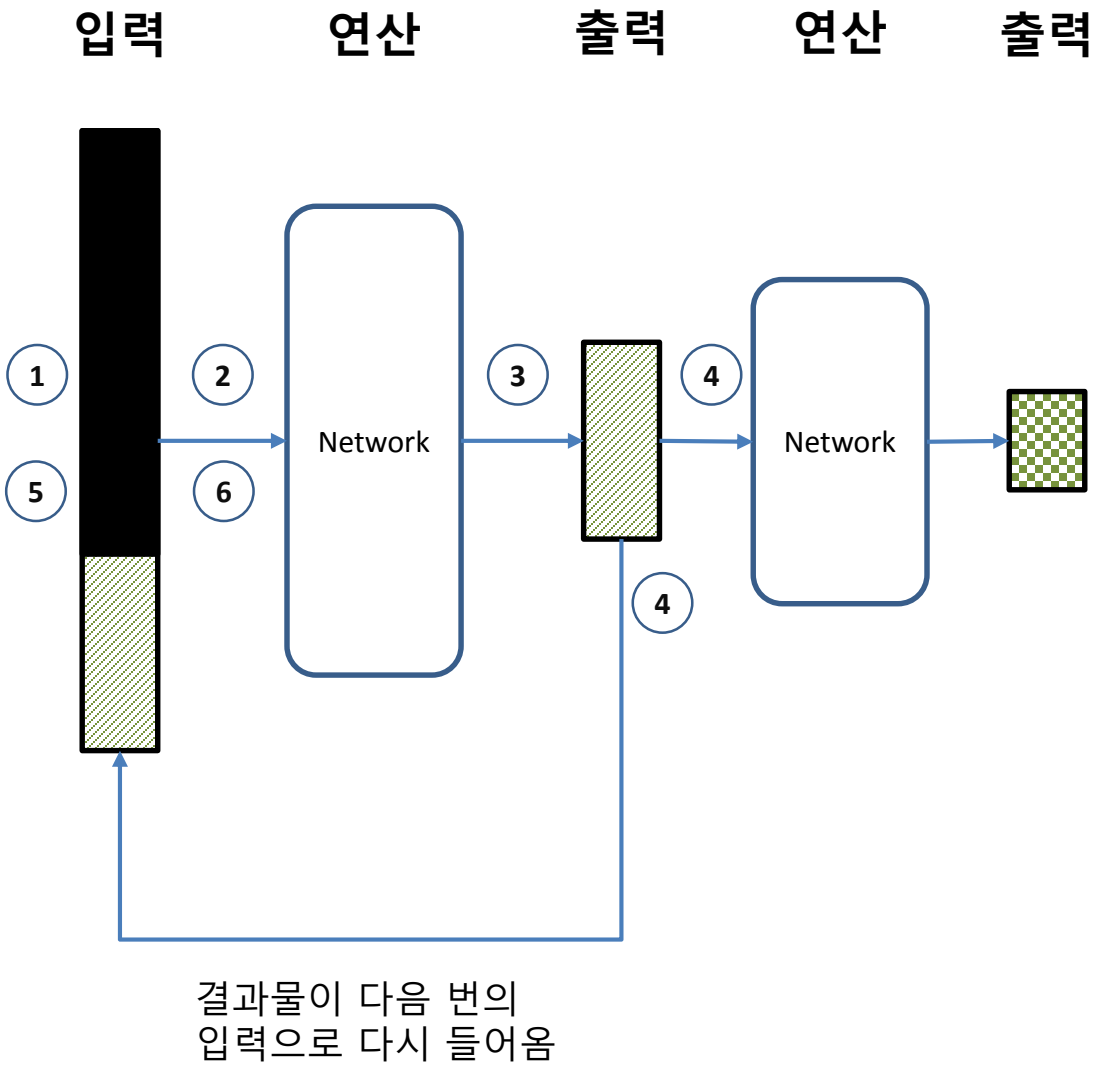


기억*이란 무엇인가?*

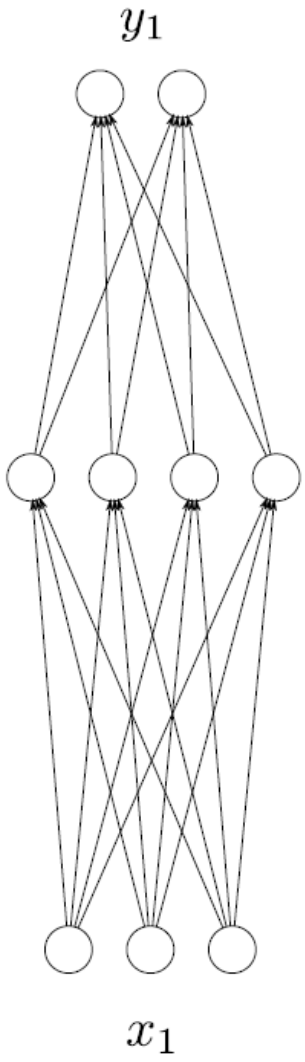
→ (무엇인지는 모르지만)
과거의 어떤 것이 현재에 영향을 미치는 것



Neural Network + Memory = Recurrent Neural Network

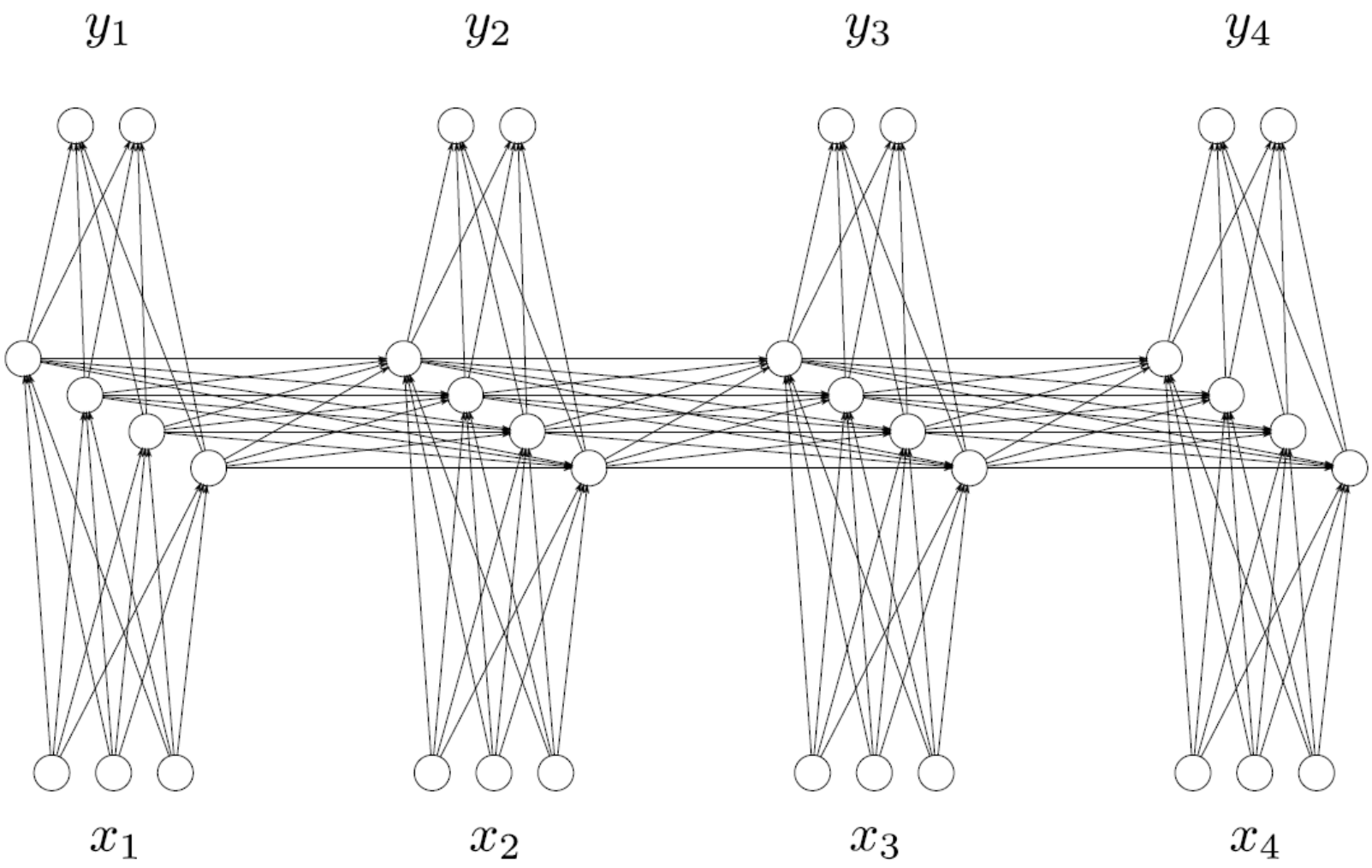


Feed forward network



http://lstm.iupr.com/

Recurrent Neural Network



http://lstm.iupr.com/

Recurrent Neural Network 문제점

- Vanishing Gradient Problem
  - 구조적으로 바로 전의 것에 영향을 받게 되어 있기 때문에 Long-Dependancy 가 있는 입력 혹은 결과가 잘 반영되지 않는다.

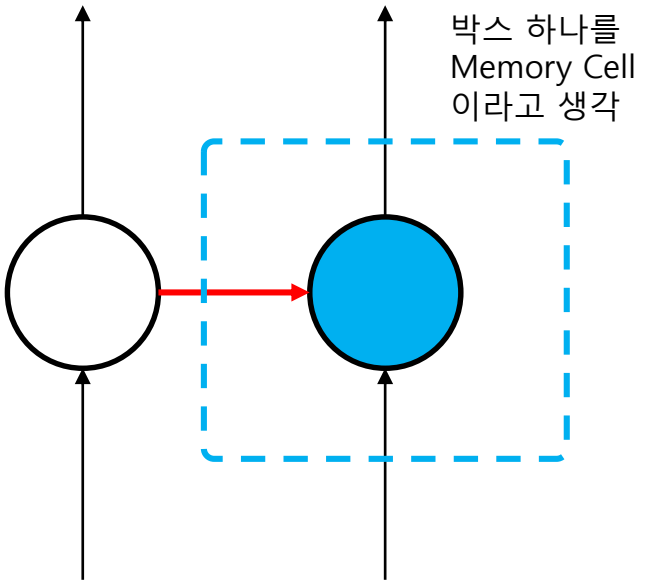
- Exploding Gradient Problem
  - Short-Dependancy 가 더 중요함에도, 멀리 있는 gradient가 너무 높게 계산되, 가까이 있는 gradient 가 충분히 반영되지 않는 경우

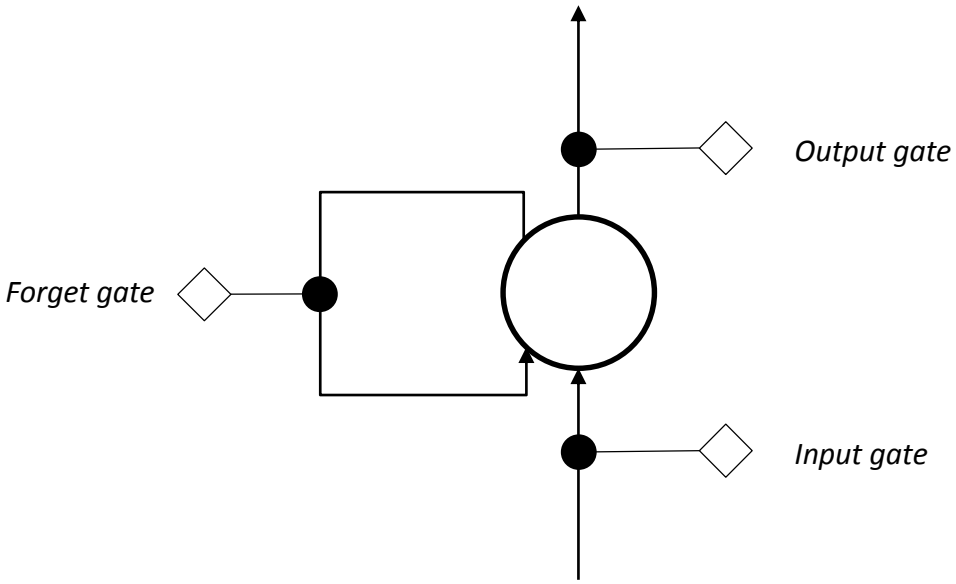
- 해결방법
  - 1990년대에 여러가지 방법론들이 나타남
    - 그 중에 가장 효과적인 Solution 으로 평가 받는 것이 LSTM

LSTM (Long Short-Term Memory)

- 멀리 있는 (Long) 정보(memory) 도 계속해서 유지될 수 있도록 하는 Neural Network 구조
- 일반적인 Back Propagation Through Time(BPTT)의 방법을 그대로 유지하면서 활용 할 수 있는 방법
- 컴퓨터에 활용되는 Memory Circuit 과 유사한 형태의 Memory 를 Neural Network 로 구현

LSTM 기본적인 Idea





neural	memory	의미
input	Write	1이면 입력 x가 들어 올수 있도록 허용(open). 0이면 Block(closed)
output	Read	1이면 의미있는 결과물로 최종 Output(open). 0이면 해당 연산 출력 안함(closed)
forget	Reset	1이면 바로 전 time 의 memory 를 유지. 0이면 reset. Keep gate

LSTM - Preservation of gradient information

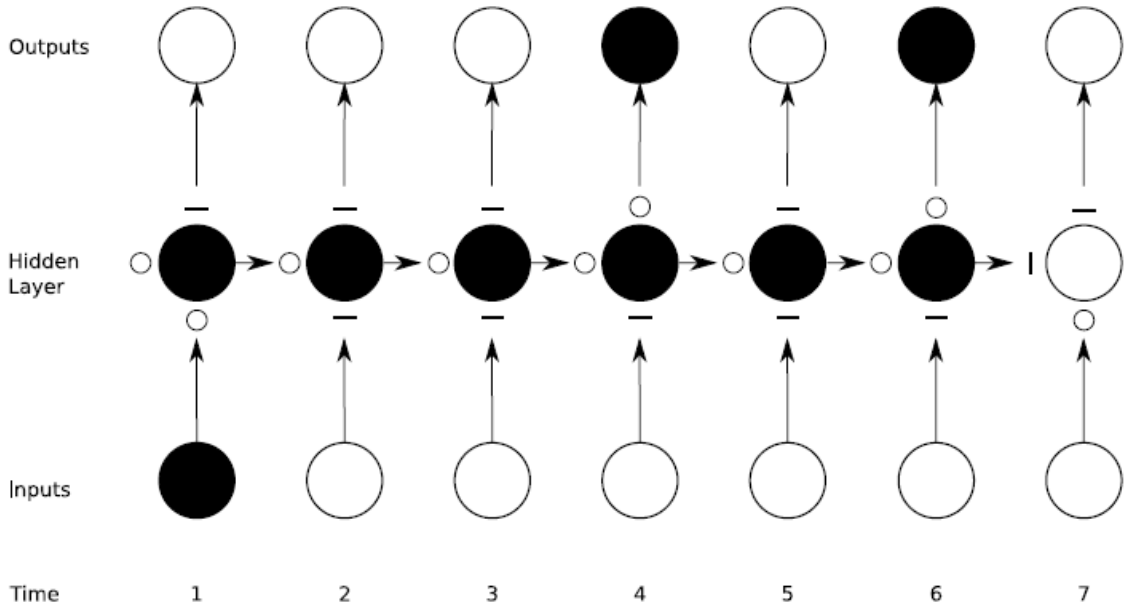
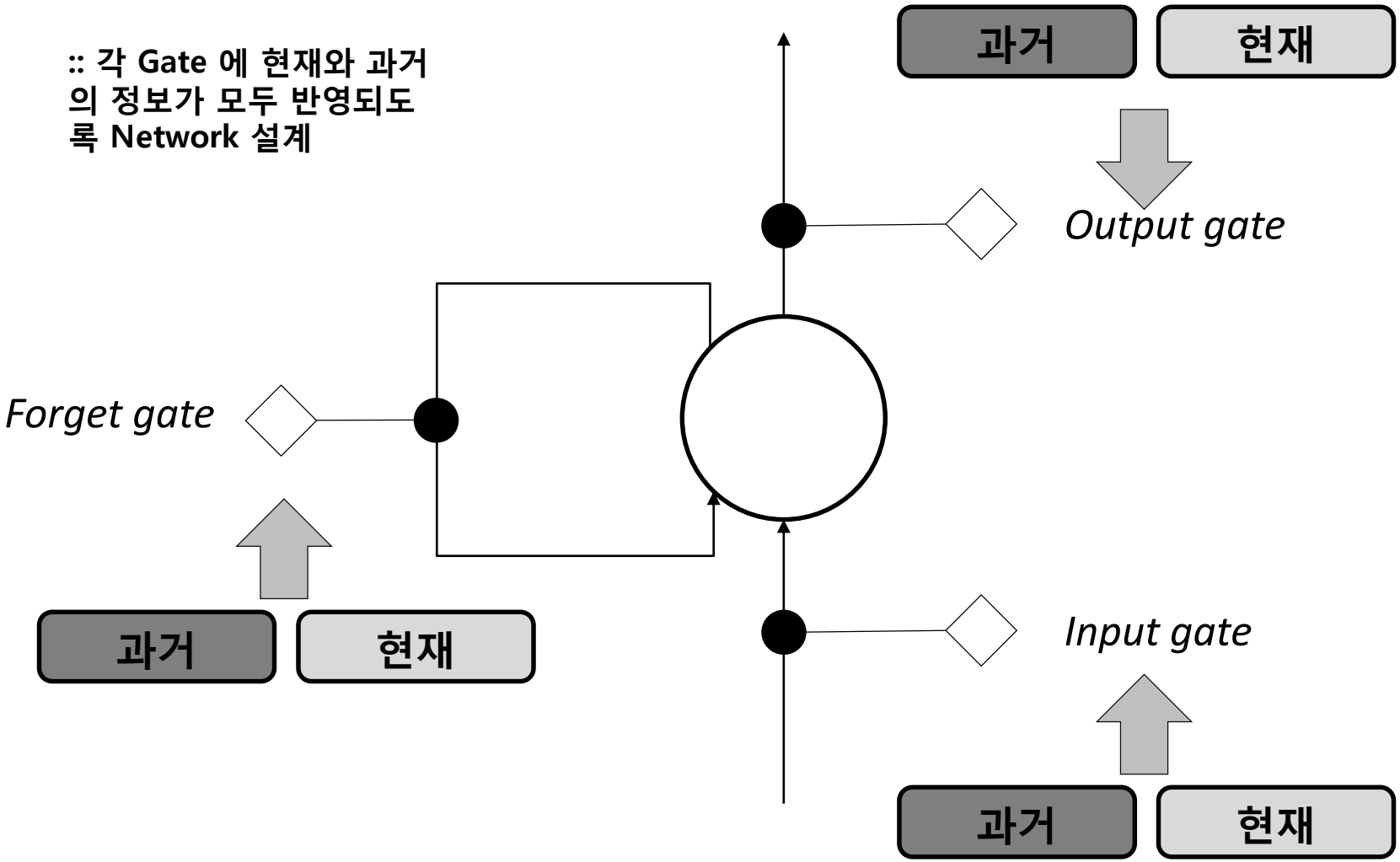


Figure 4.3: **Preservation of gradient information by LSTM.** As in Figure 4.1 the shading of the nodes indicates their sensitivity to the input unit at time one. The state of the input, forget, and output gate states are displayed below, to the left and above the hidden layer node, which corresponds to a single memory cell. For simplicity, the gates are either entirely open ('O') or closed ('—'). The memory cell 'remembers' the first input as long as the forget gate is open and the input gate is closed, and the sensitivity of the output layer can be switched on and off by the output gate without affecting the cell.

http://www.cs.toronto.edu/~graves/phd.pdf

LSTM - Idea

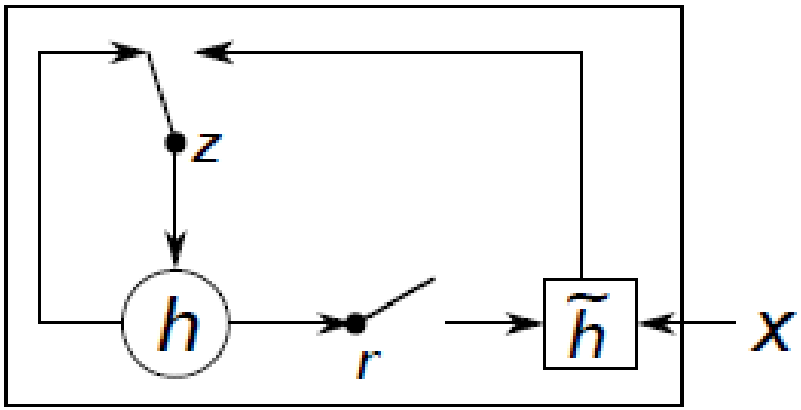
:: 각 Gate 에 현재와 과거의 정보가 모두 반영되도록 Network 설계



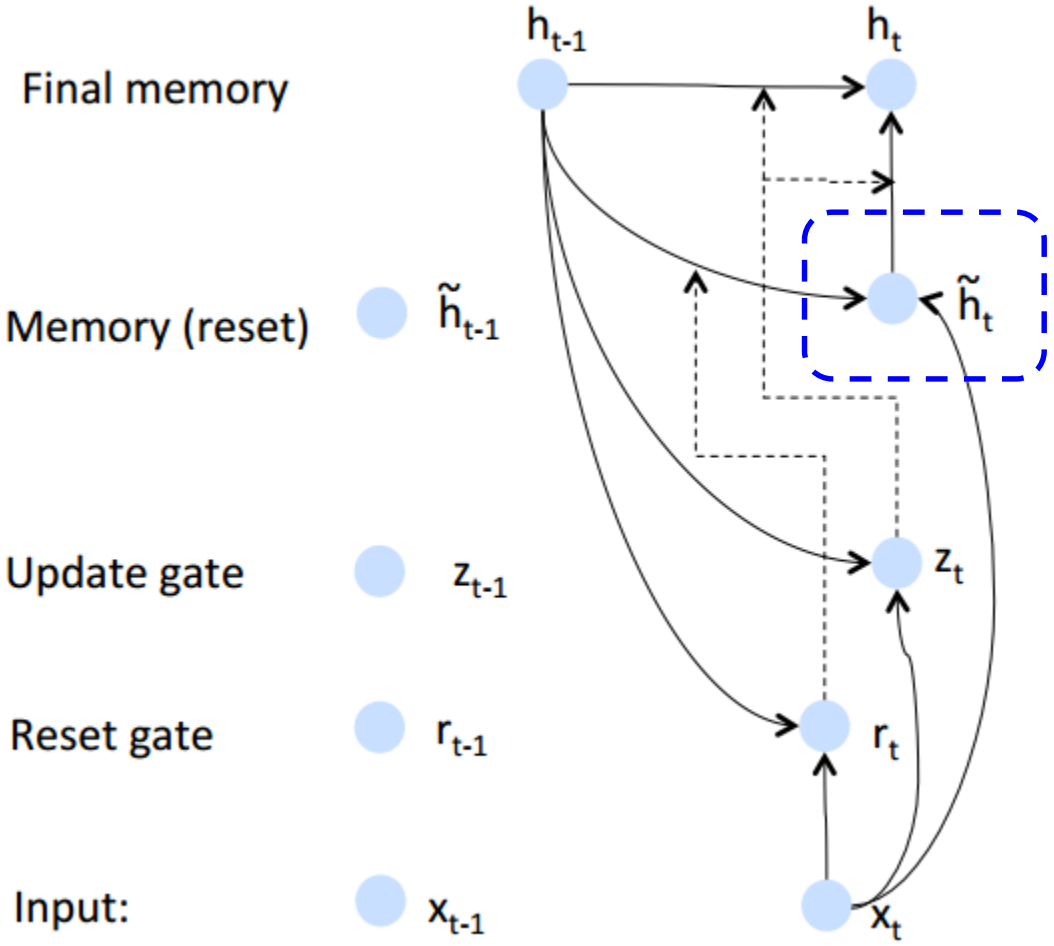
Gated Recurrent Unit

- Update Gate 를 두어서 현재 Time 의 Hidden state 를 계산할 때 update gate 의 영향을 받도록 함
- Reset Gate 를 두어서 현재 Memory 를 Reset 할지 안할지를 결정(like LSTM)

- Cho et al. 2014



Gated Recurrent Unit



$$z_t = \sigma\left(W^{(z)}x_t + U^{(z)}h_{t-1}\right)$$

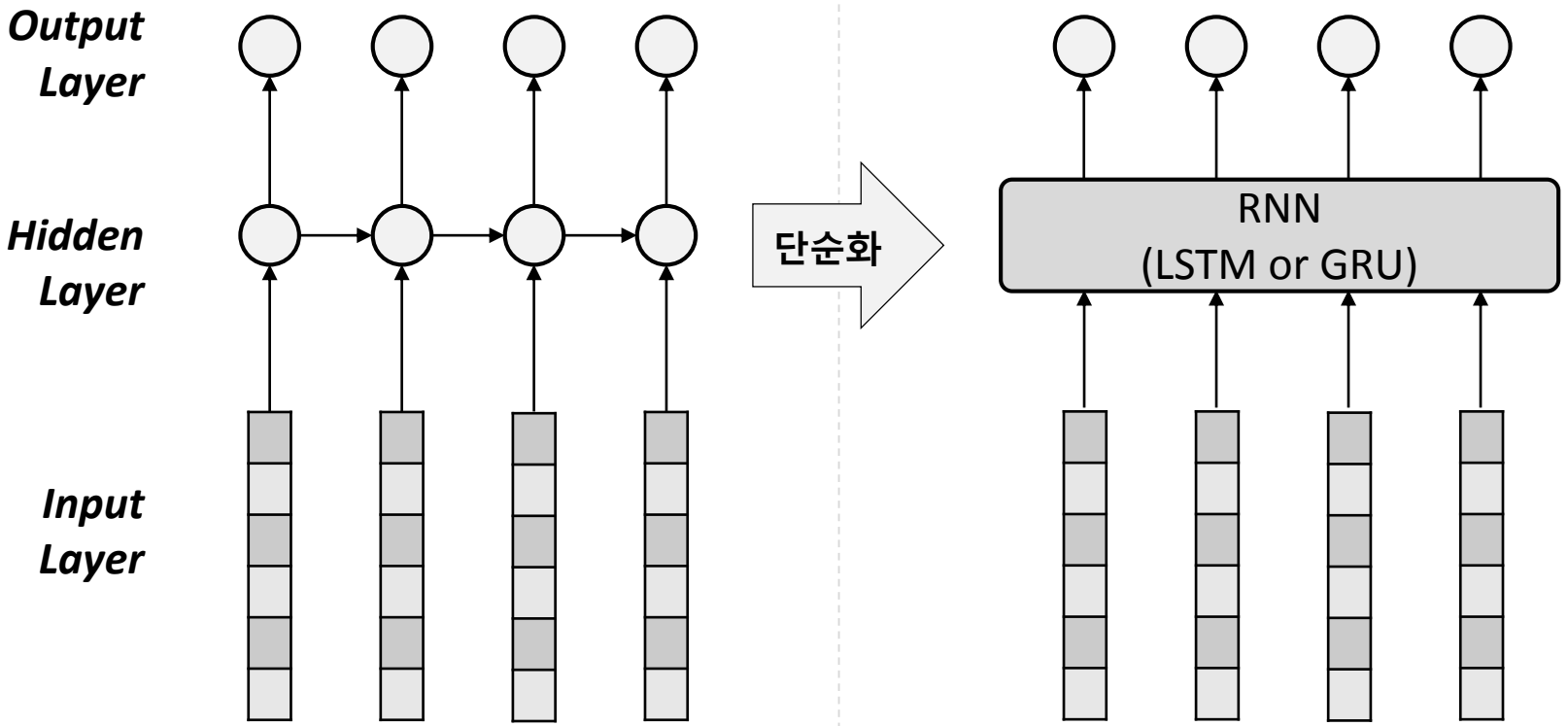
$$r_t = \sigma\left(W^{(r)}x_t + U^{(r)}h_{t-1}\right)$$

$$\tilde{h}_t = \tanh\left(Wx_t + r_t \circ Uh_{t-1}\right)$$

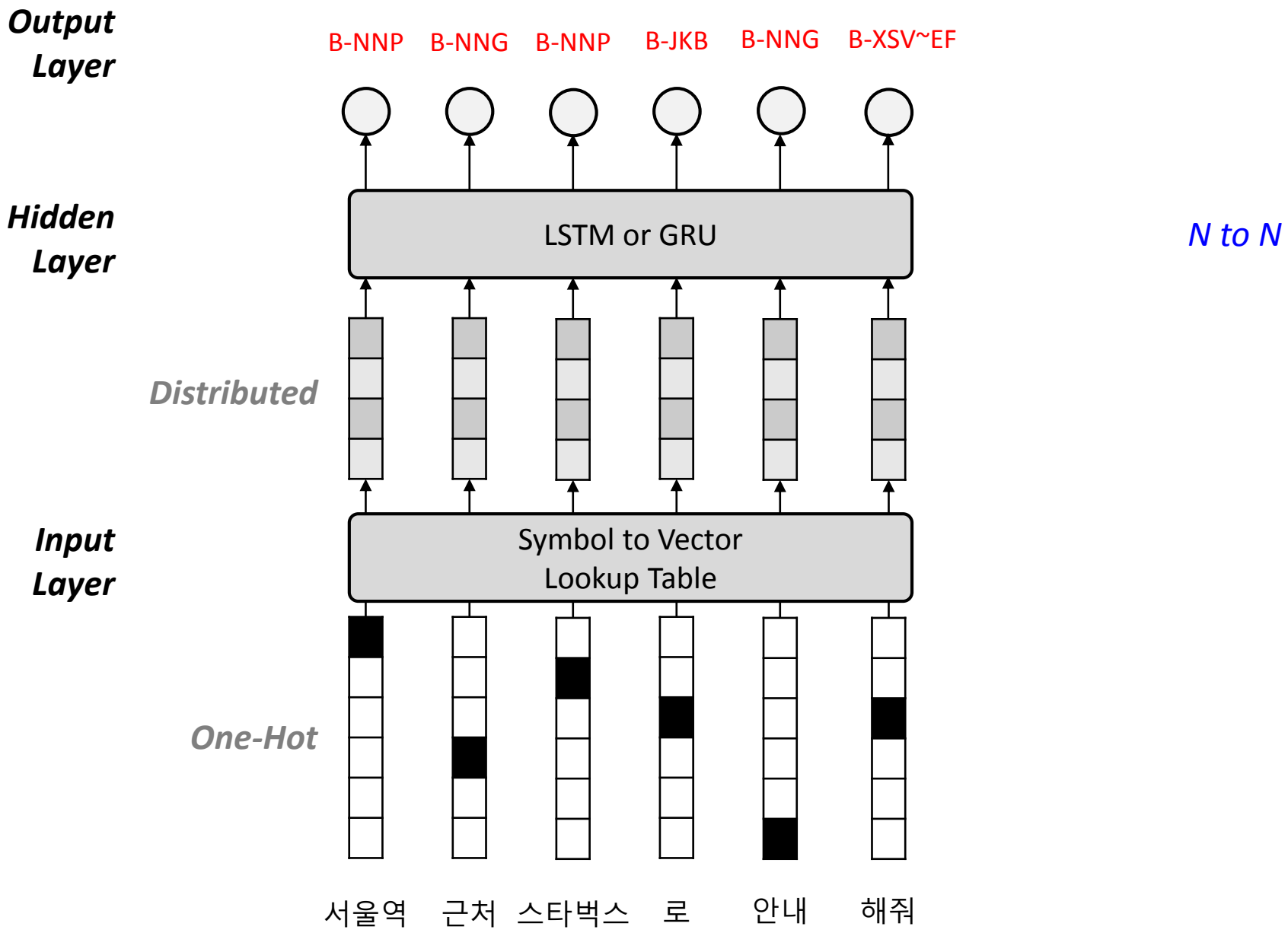
$$h_t = z_t \circ h_{t-1} + (1 - z_t) \circ \tilde{h}_t$$

http://cs224d.stanford.edu/lectures/CS224d-Lecture8.pdf

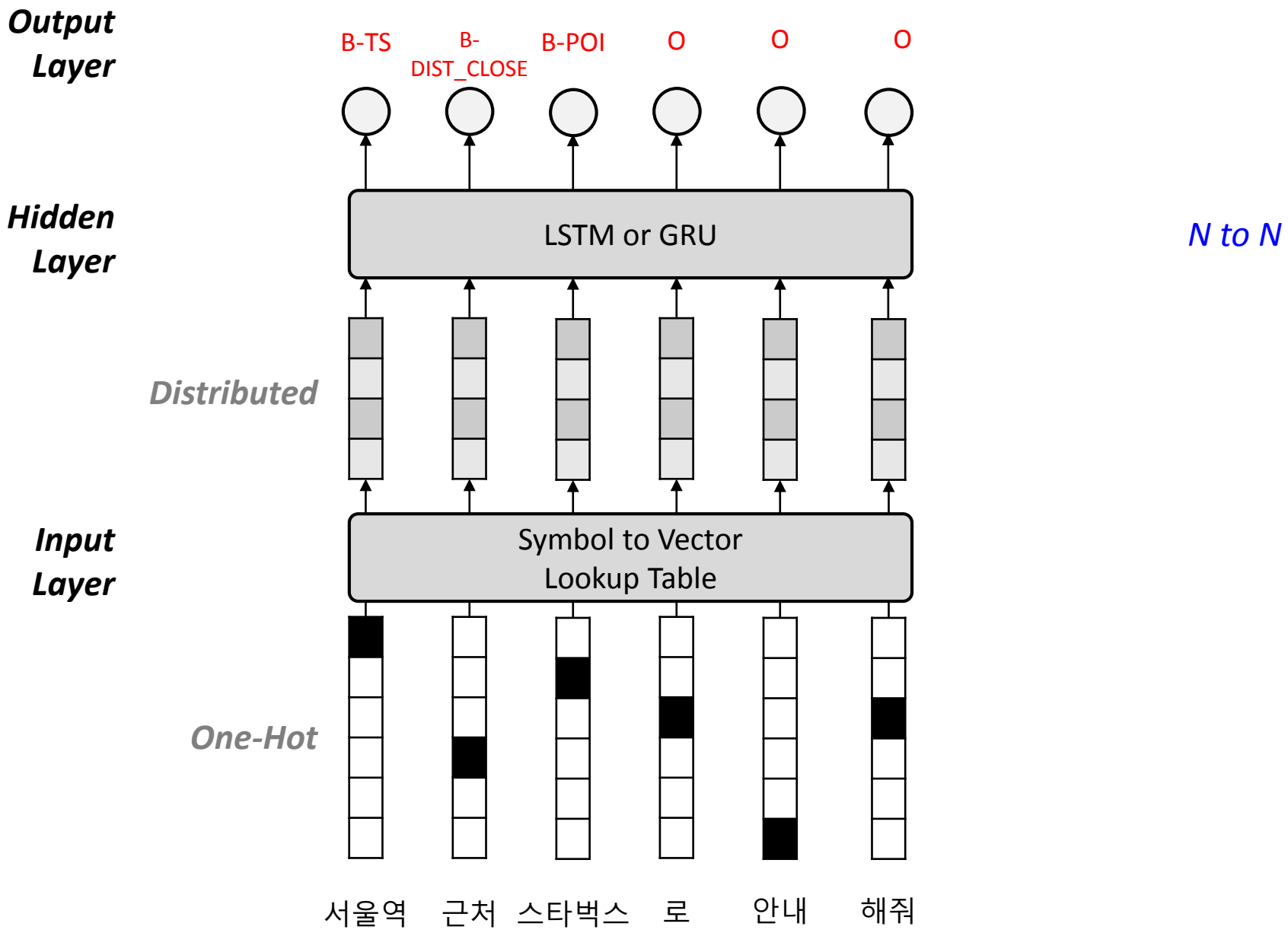
Sequential Modeling @ Deep Learning



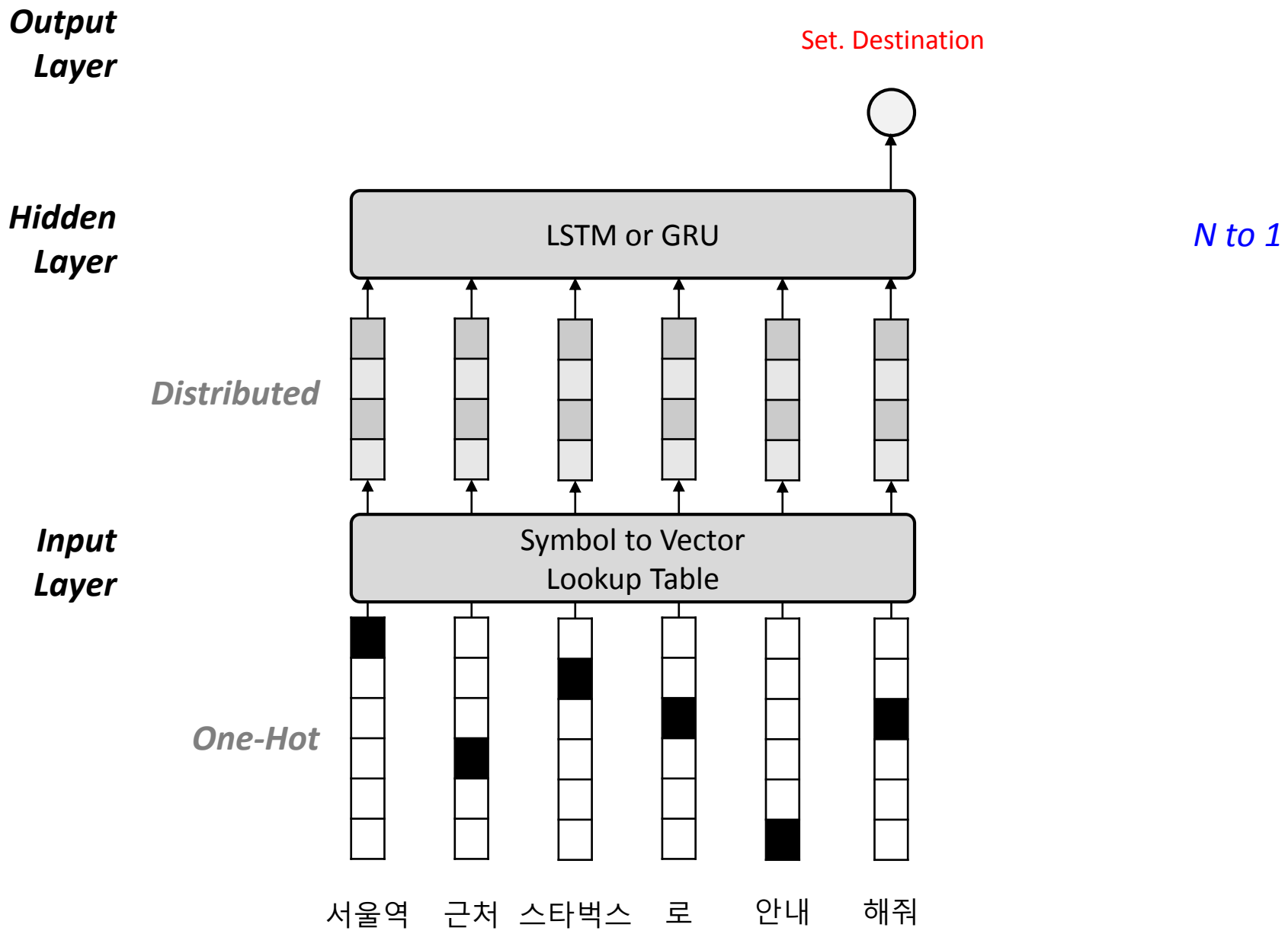
Sequence Modeling for POS Tagging



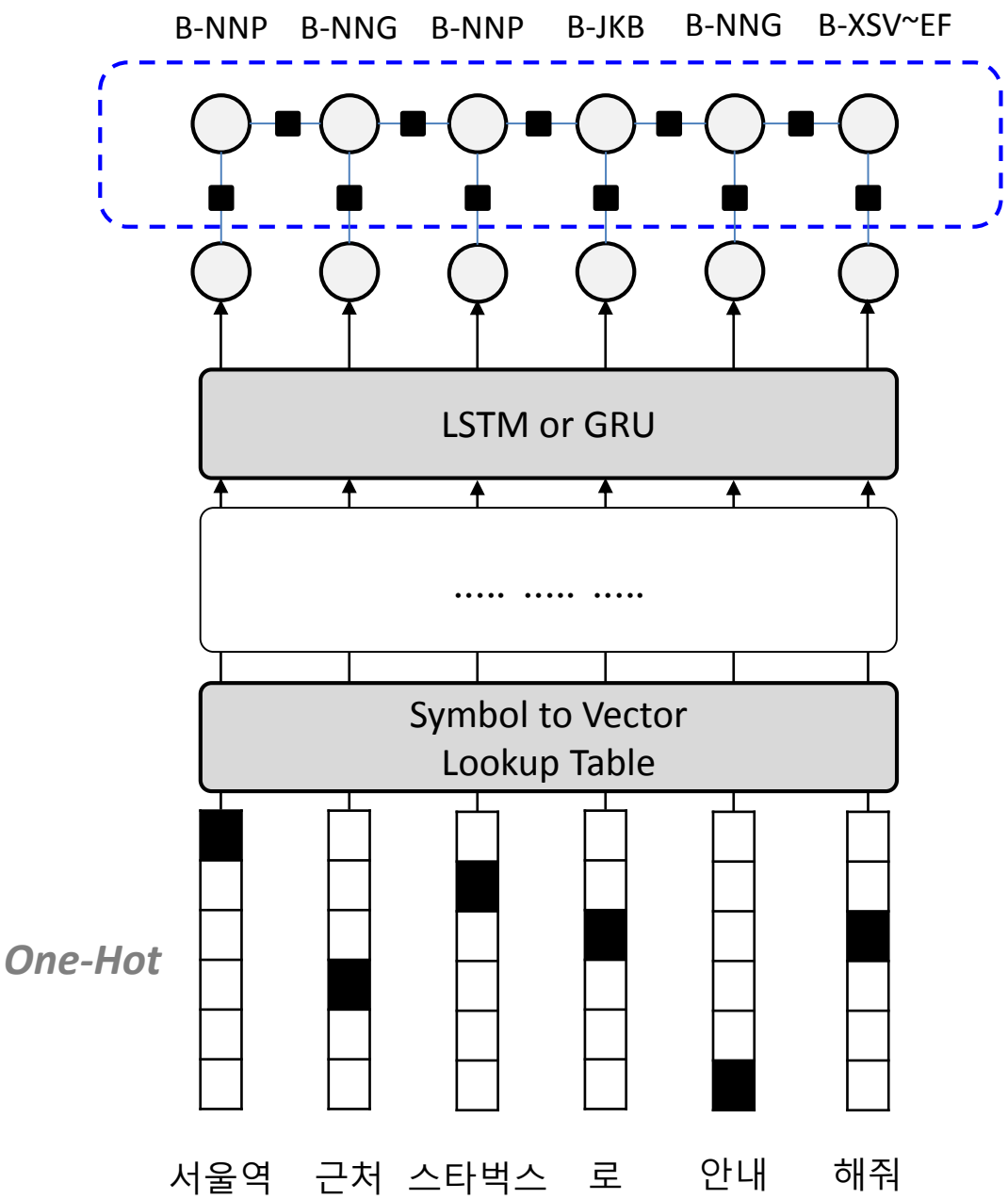
Sequence Modeling for Domain Entity Tagging



Sequence Modeling for Intention Analysis



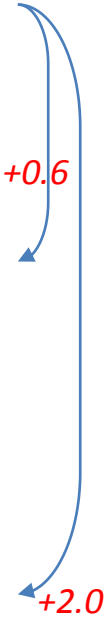
Hidden Layer Sequence Modeling + Output Sequence Modeling



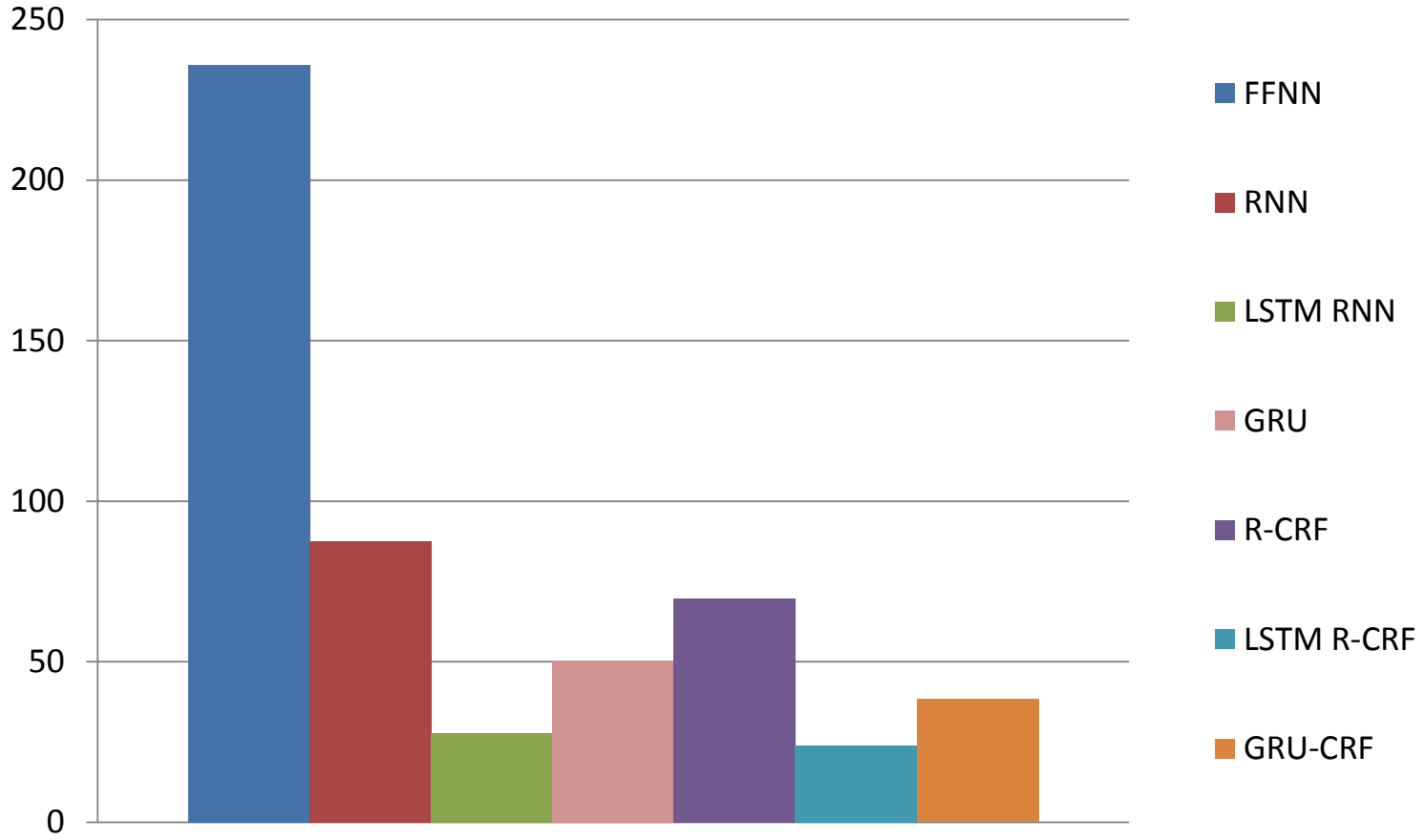
기본 RNN 구조에 output layer 에 CRF 를 붙여서 Sequential Output에 최적화 시켜 모델링 할 수도 있음

Performance

영어 개체명 인식 (CoNLL03 data set)	F1(dev)	F1(test)
Structural SVM (baseline + Word embedding feature) – no Deep Learning	-	85.58
SENNA (Collobert)	-	89.59
FFNN (Sigm + Dropout + Word embedding)	91.58	87.35
RNN (Sigm + Dropout + Word embedding)	91.83	88.09
LSTM RNN (Sigm + Dropout + Word embedding)	91.77	87.73
GRU RNN (Sigm + Dropout + Word embedding)	92.01	87.96
CNN+CRF (Sigm + Dropout + Word embedding)	93.09	88.69
RNN+CRF (Sigm + Dropout + Word embedding)	93.23	88.76
LSTM+CRF (Sigm + Dropout + Word embedding)	93.82	90.12
GRU+CRF (Sigm + Dropout + Word embedding)	93.67	89.98



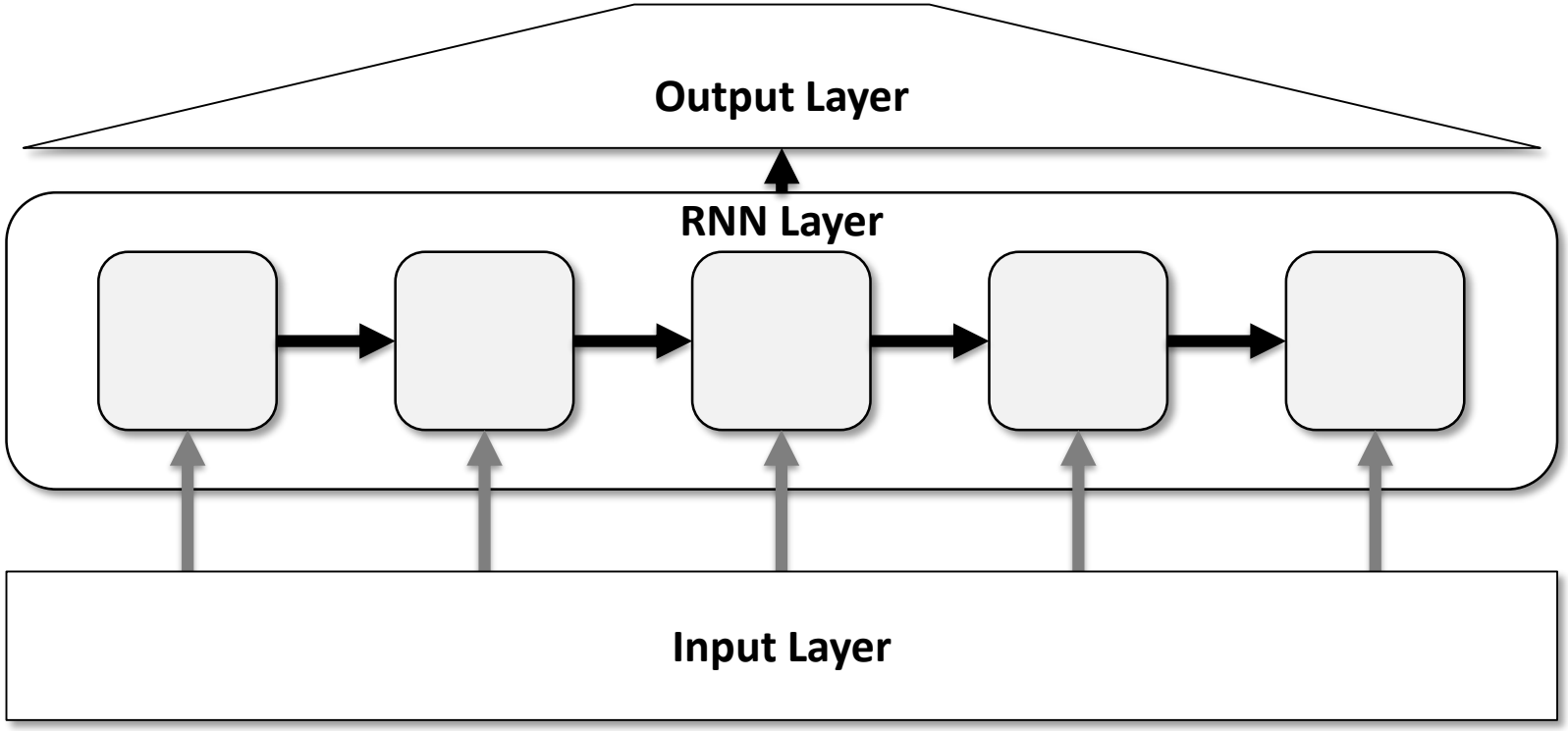
학습속도


250
200
150
100
50
0
FFNN
RNN
LSTM RNN
GRU
R-CRF
LSTM R-CRF
GRU-CRF

Recent Movement

ENCODING – DECODING APPROACH

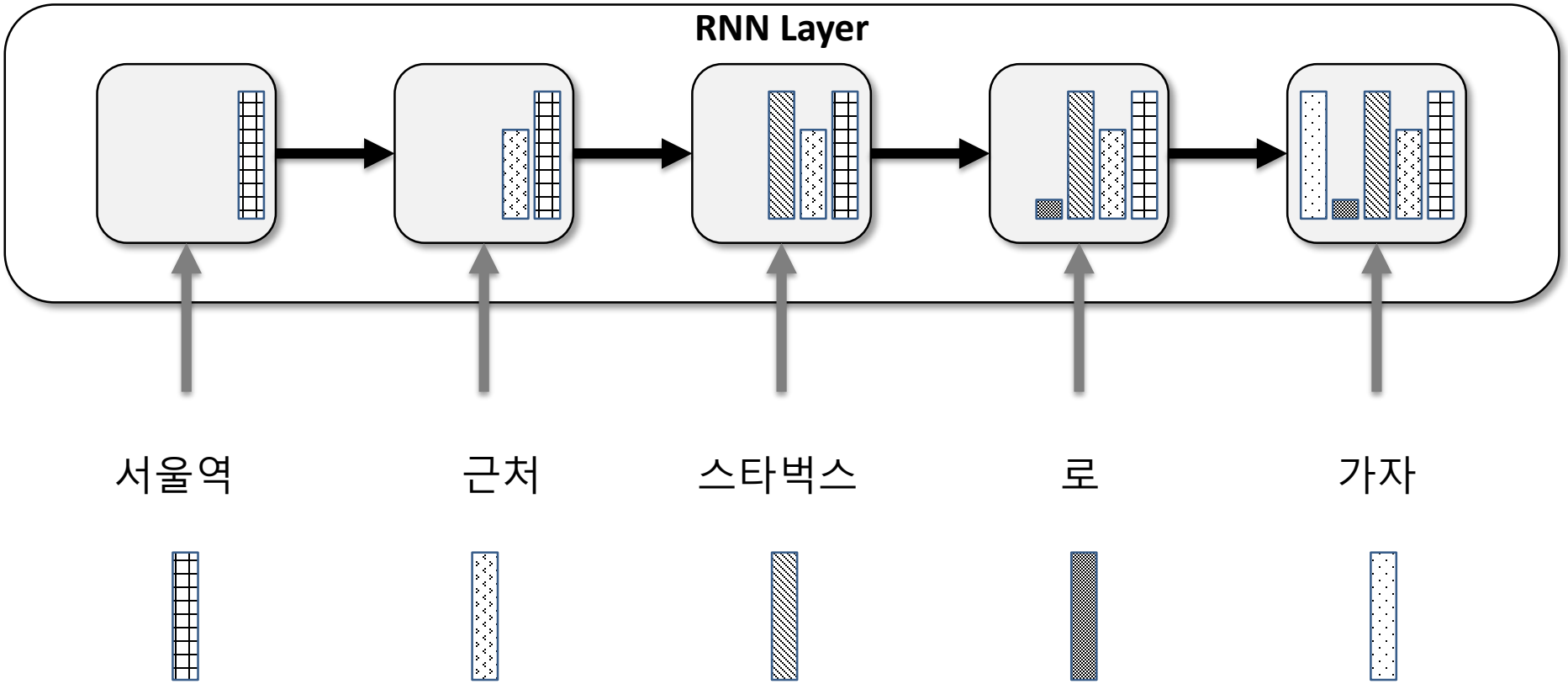
Recurrent Neural Network - Review



Output 이 잘나오도록 하는 정보를 RNN Layer 에 기억하게 됨

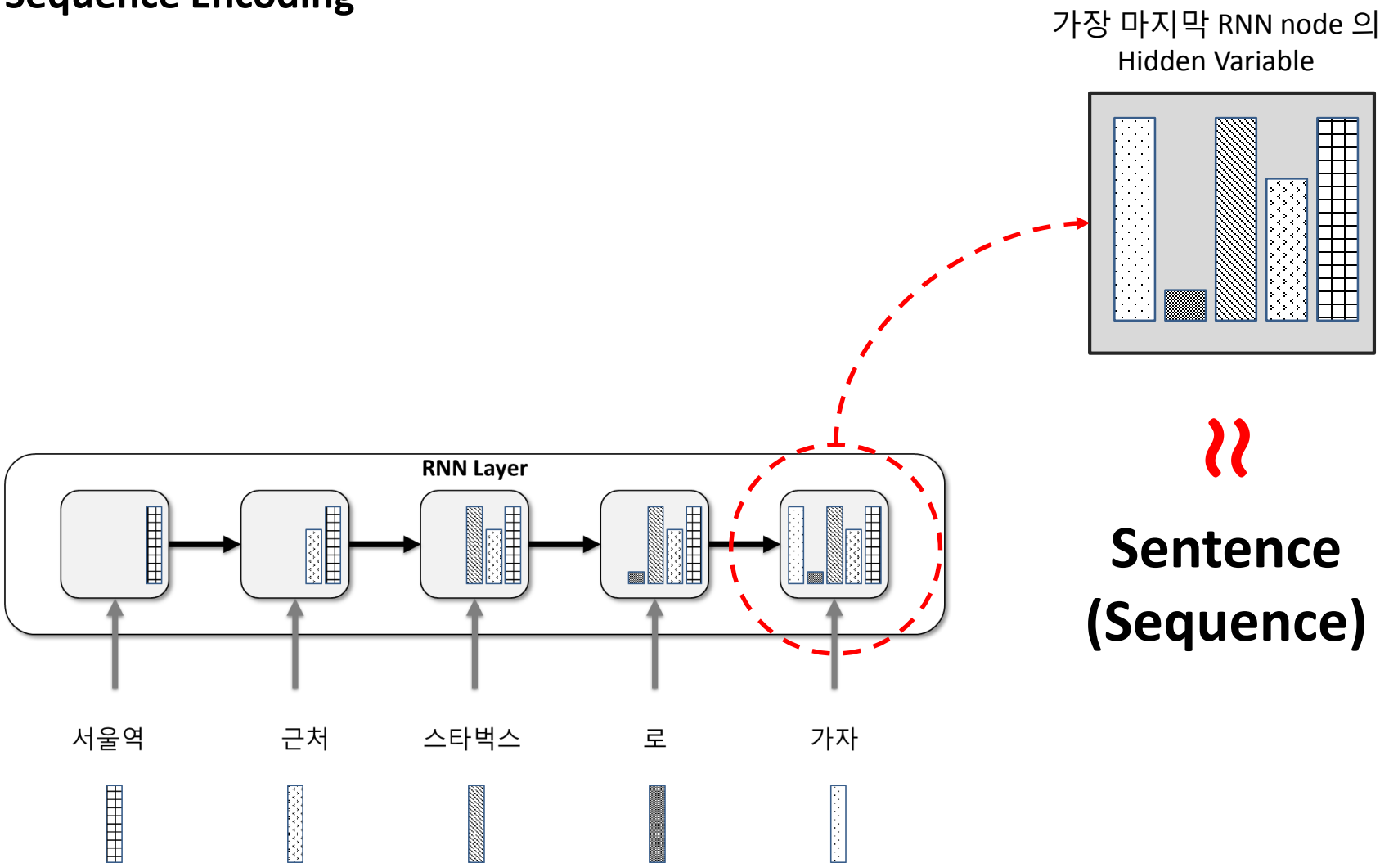
Recurrent Neural Network - Review

Output 이 잘나오도록 하는 정보를 RNN Layer 에 기억하게 됨



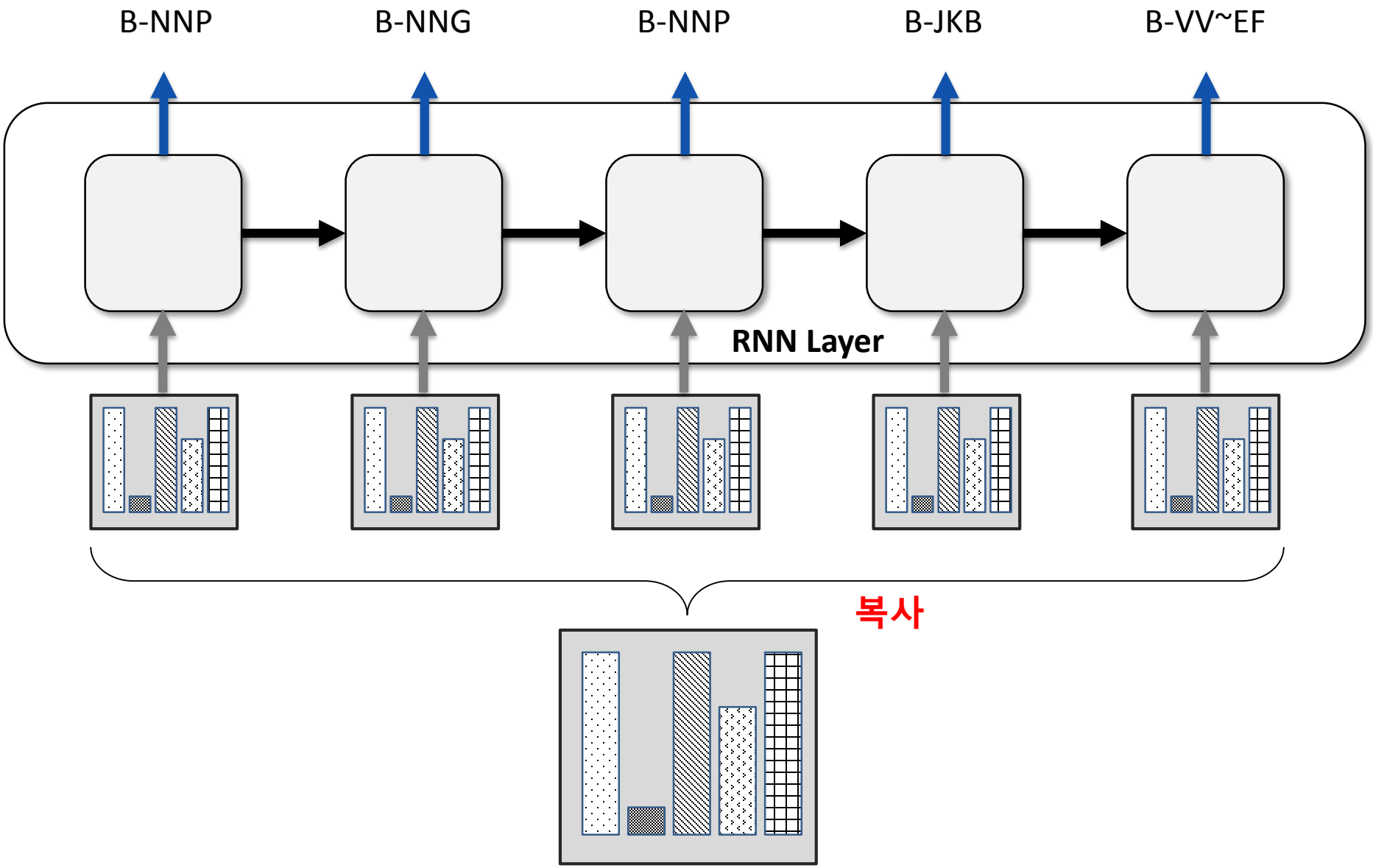
:: 이해의 편의를 위한 도식화.
실제로는 input dimension 과 Hidden layer dimension 이 다름

Sequence Encoding

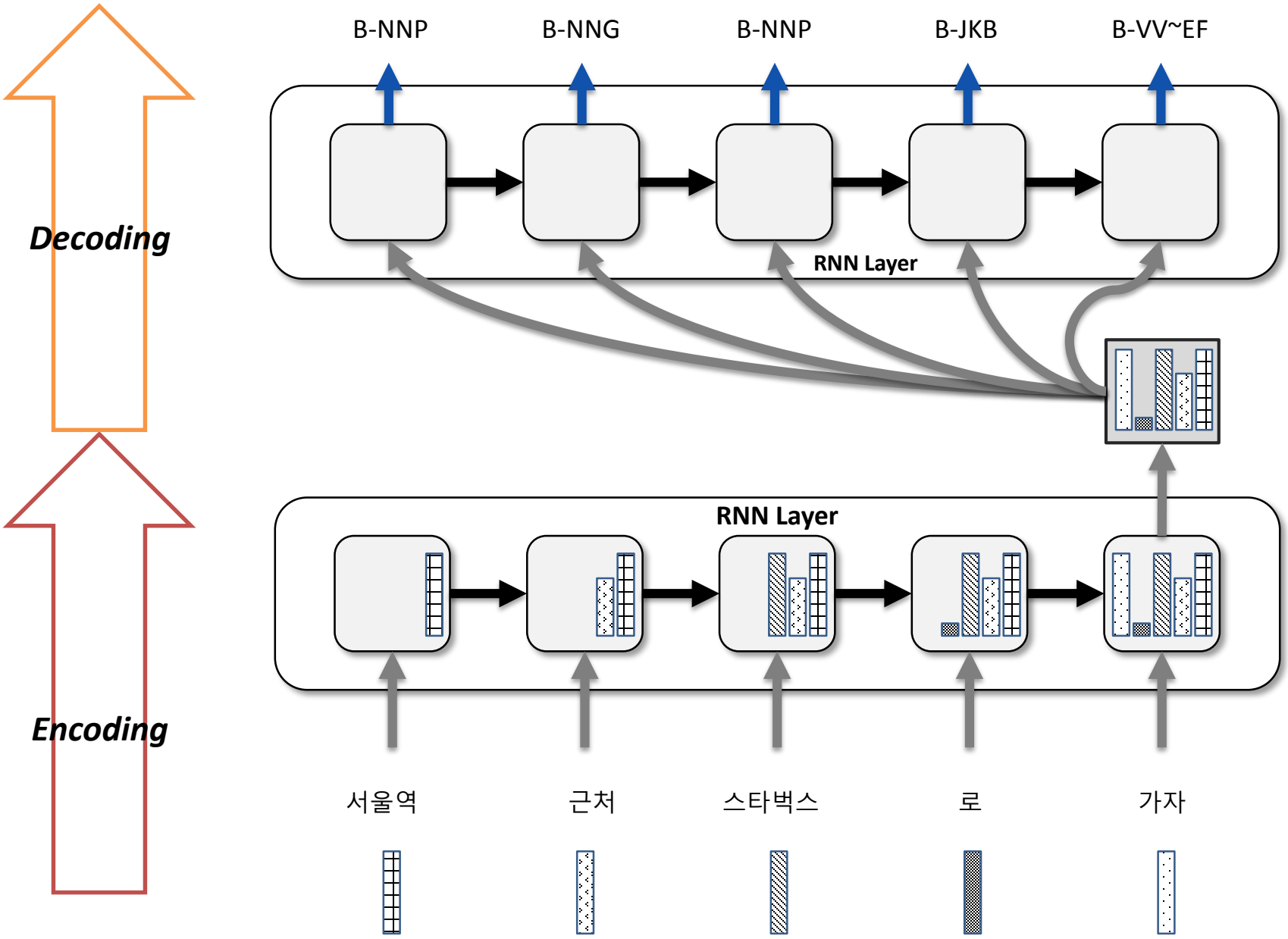


Idea : RNN 에 누적된 정보가 결국 Sequence 의 Vector Form 일 것이다.

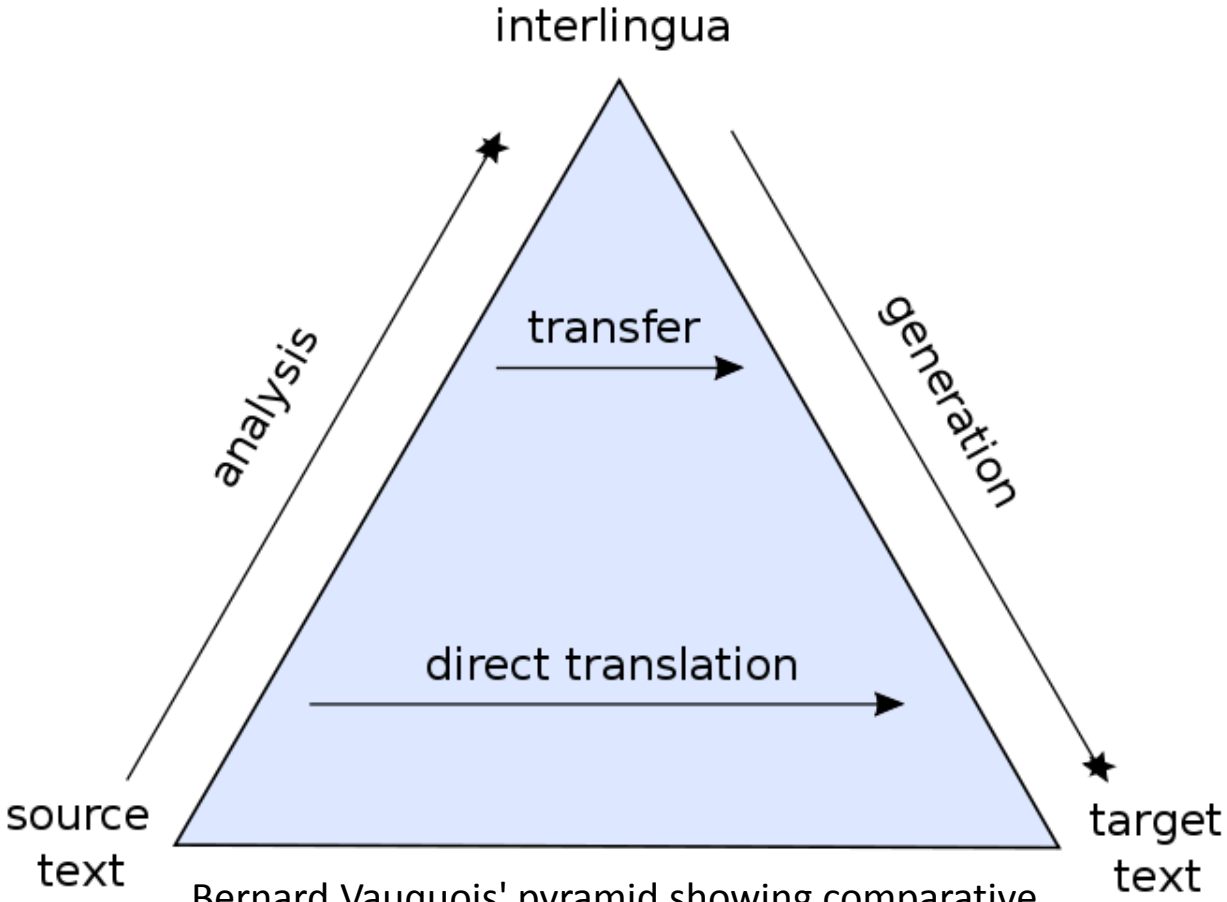
Sequence Decoding



Sequence Encoding-Decoding Approach

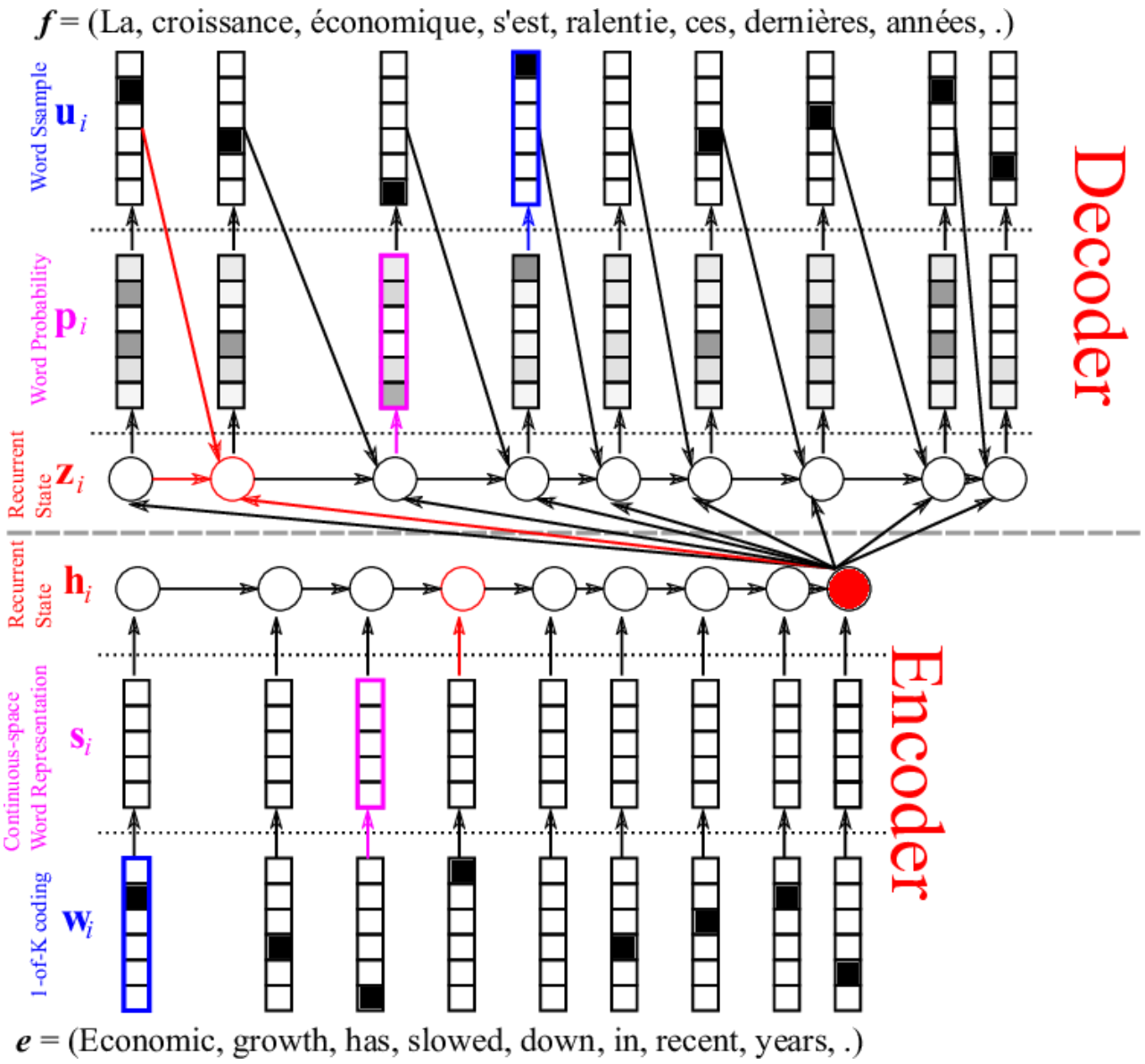


Neural Machine Translation



Bernard Vauquois' pyramid showing comparative depths of intermediary representation, interlingual machine translation at the peak, followed by transfer-based, then direct translation.
[http://en.wikipedia.org/wiki/Machine_translation]

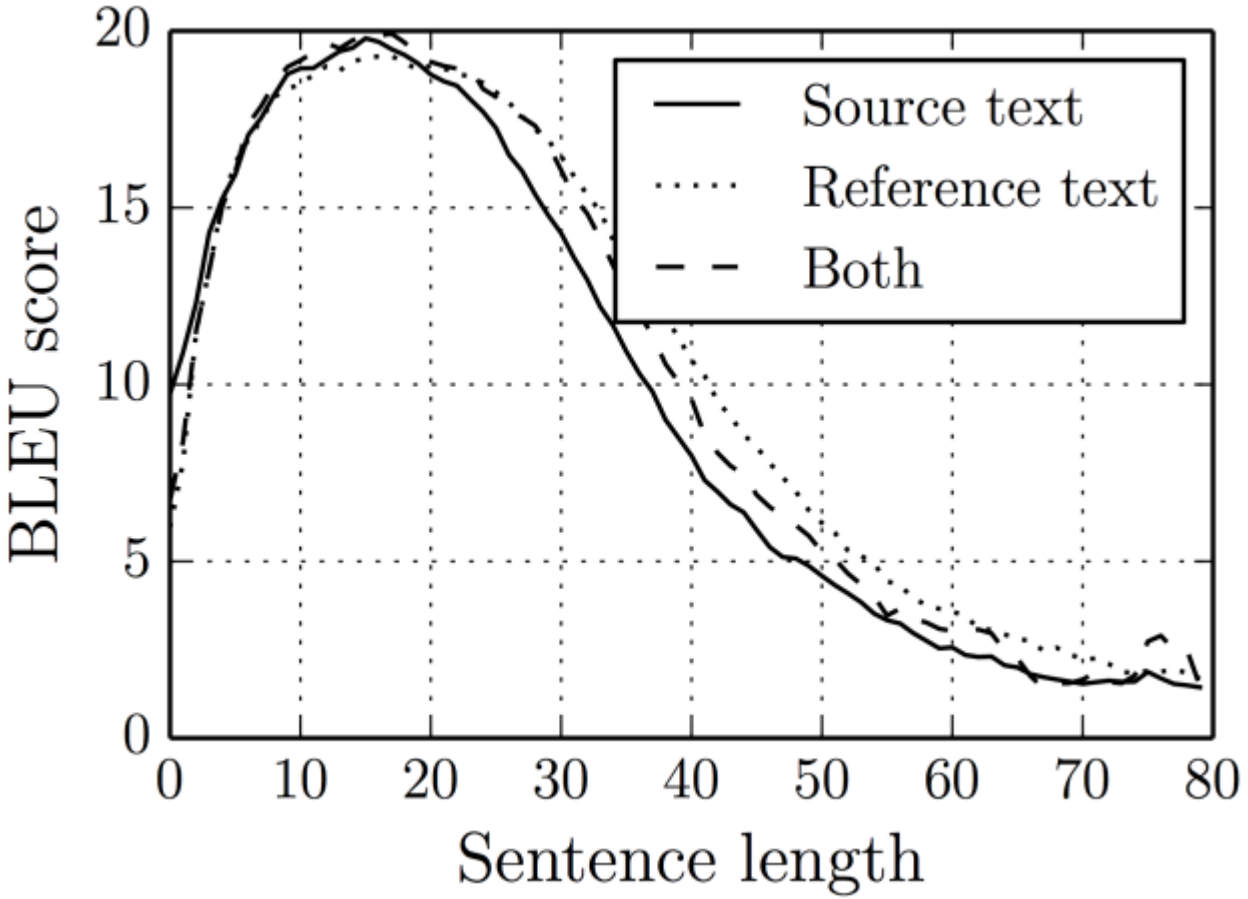
RNN Encoder-Decoder for Machine Translation



Cho et al. (2014)

http://devblogs.nvidia.com/parallelforall/introduction-neural-machine-translation-gpus-part-2/

Limitation - RNN Encoder-Decoder Approach



문장이 길어지면, 제한된 hidden variable 에 정보를 충분히 담지 못하게 되어, 긴 문장의 경우 번역이 잘되지 않는다.

http://devblogs.nvidia.com/parallelforall/introduction-neural-machine-translation-gpus-part-3/

Attention Modeling

:: 한국어 → 영어 번역

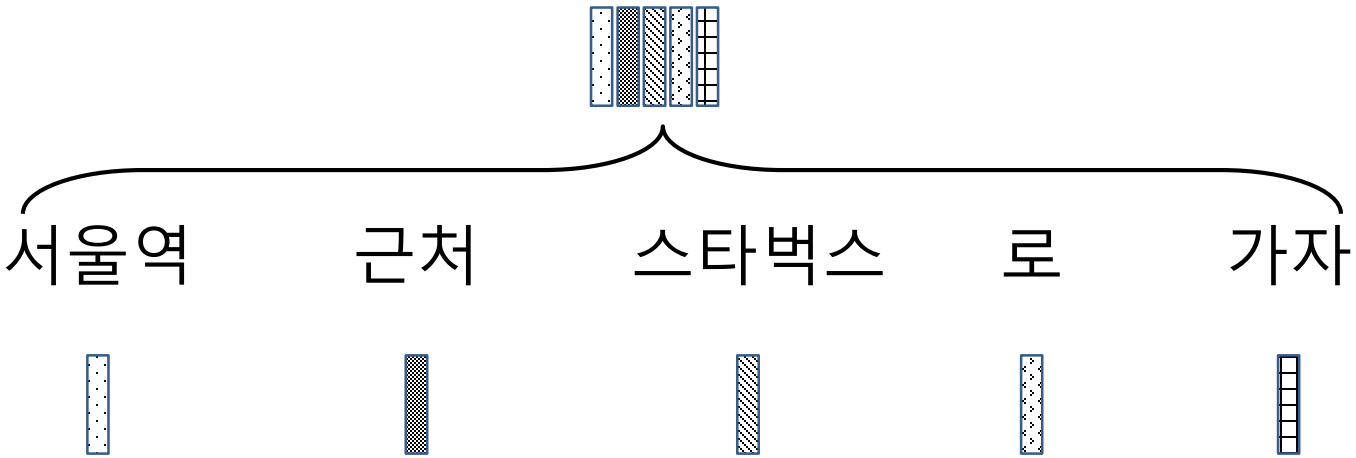
Let’s go to Starbucks near Seoul station

서울역 근처 스타벅스 로 가자

Attention Modeling

Encoding

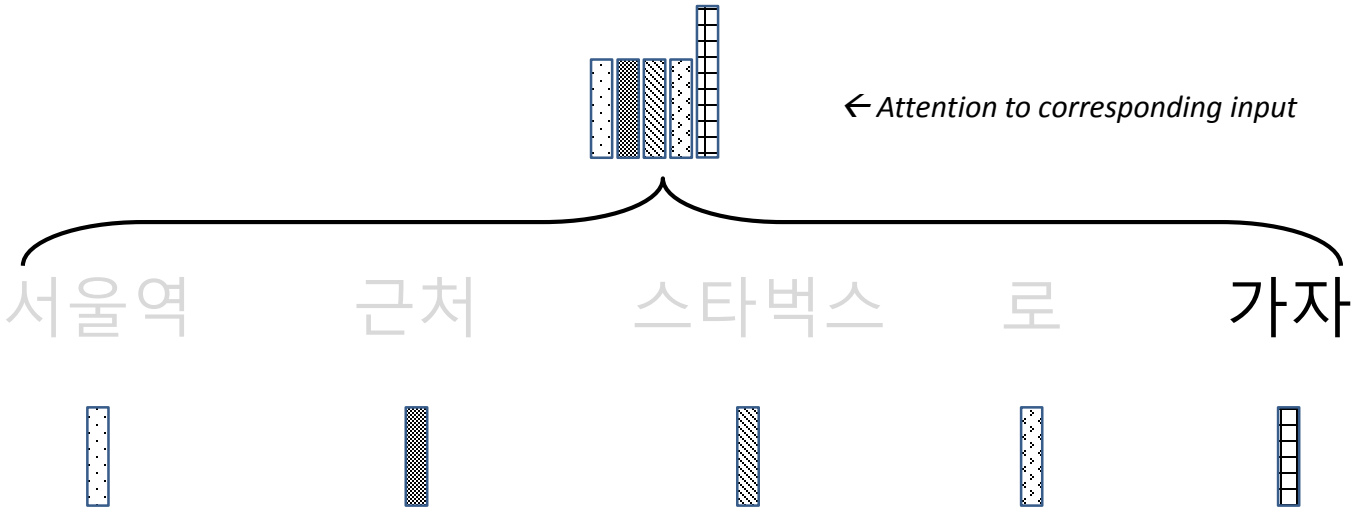
Let’s go to Starbucks near Seoul station



Attention Modeling

Decoding

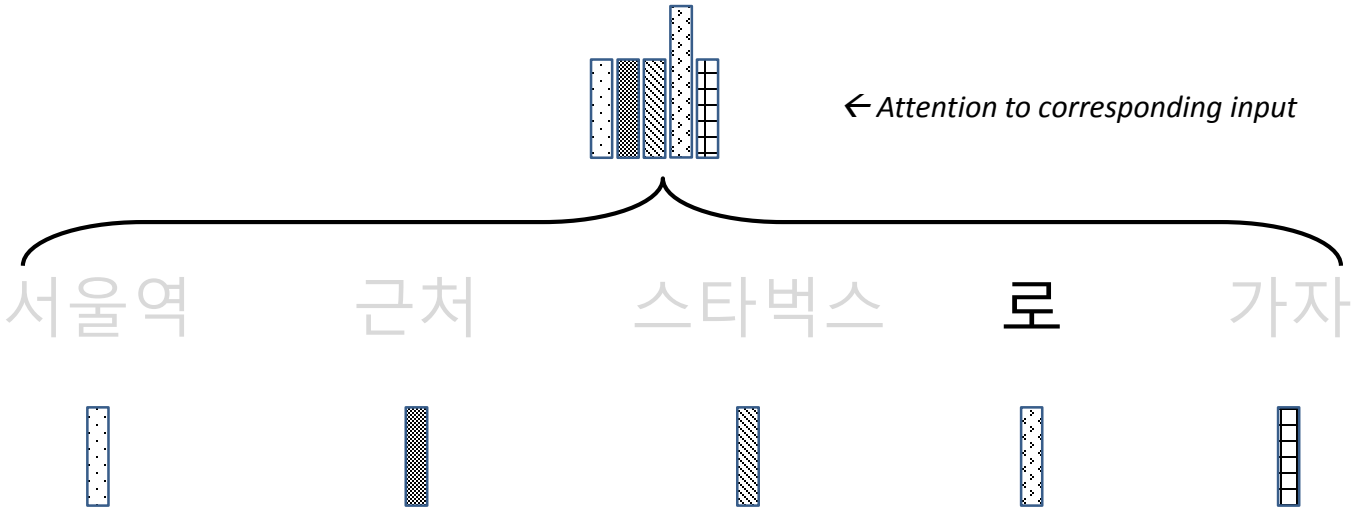
Let’s go to Starbucks near Seoul station



Attention Modeling

Decoding

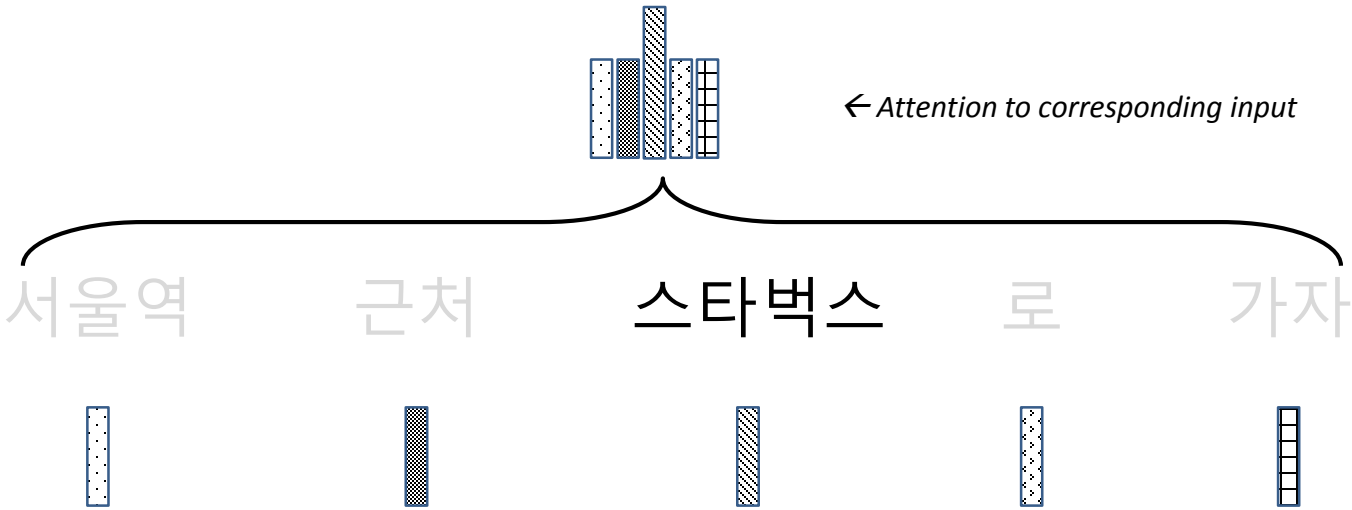
Let’s go to Starbucks near Seoul station



Attention Modeling

Decoding

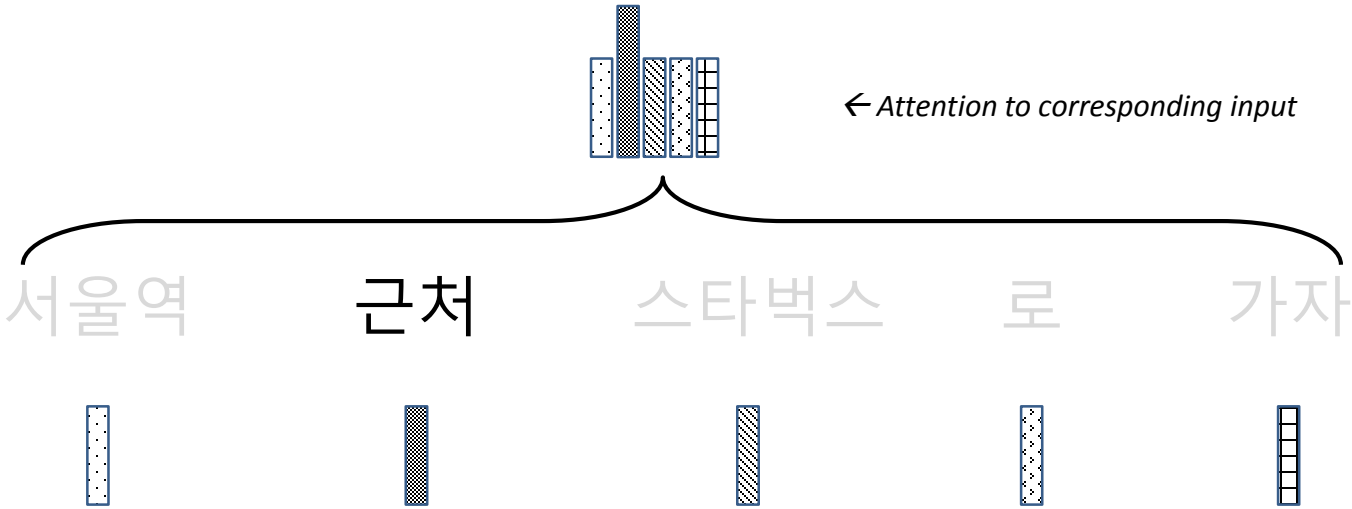
Let’s go to Starbucks near Seoul station



Attention Modeling

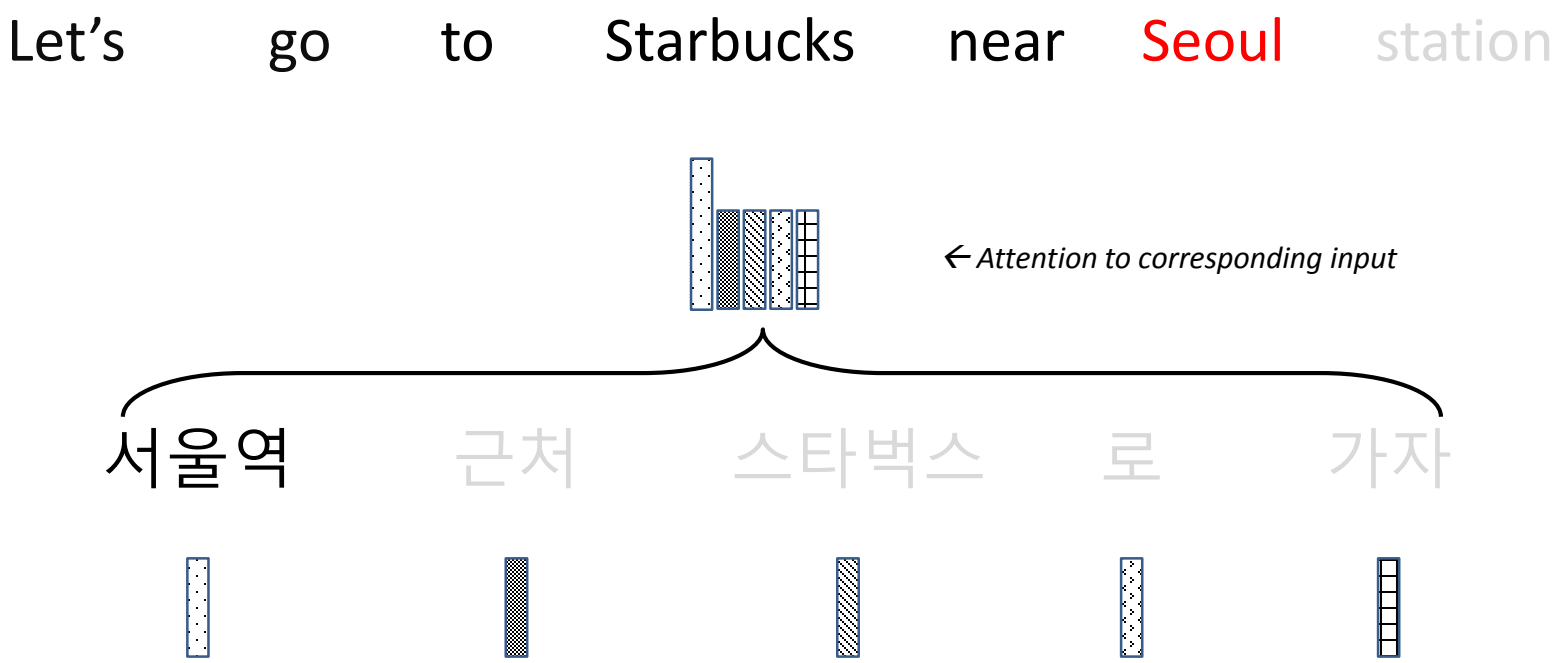
Decoding

Let’s go to Starbucks near Seoul station



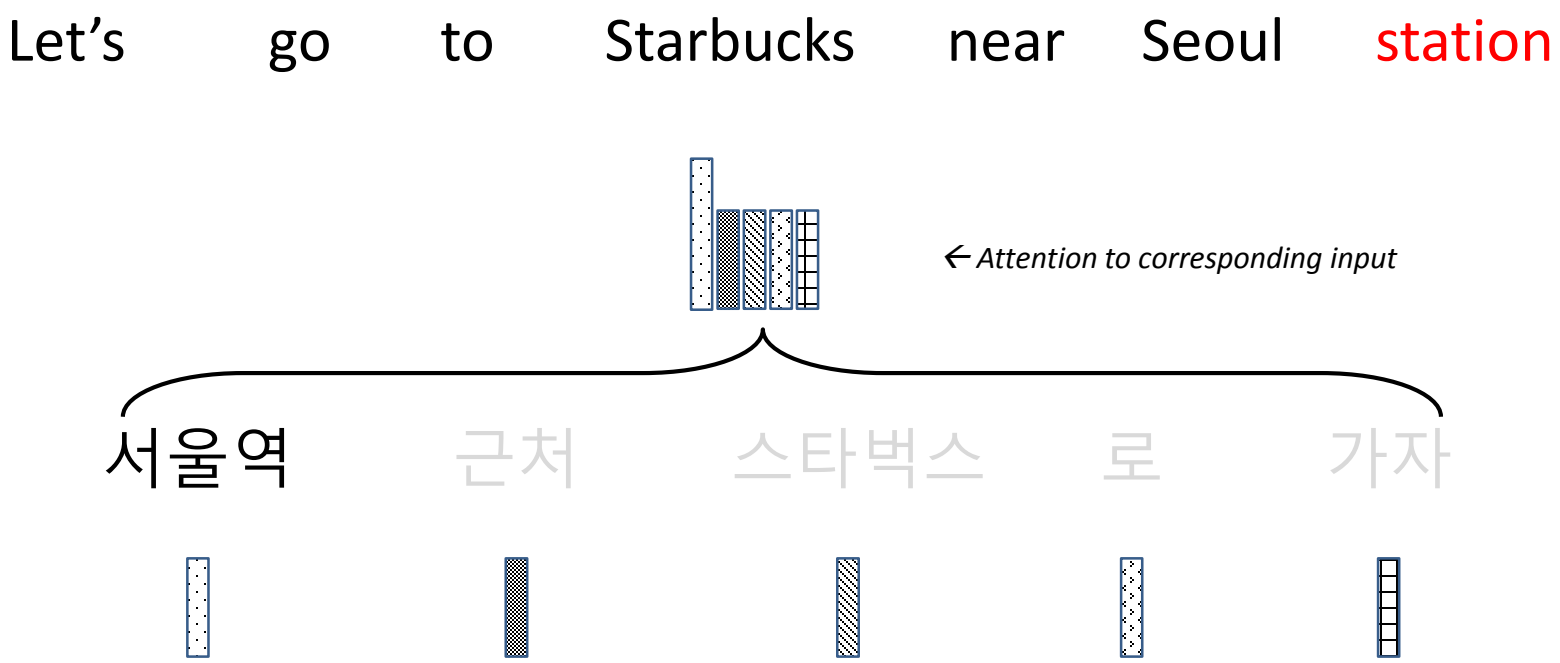
Attention Modeling

Decoding



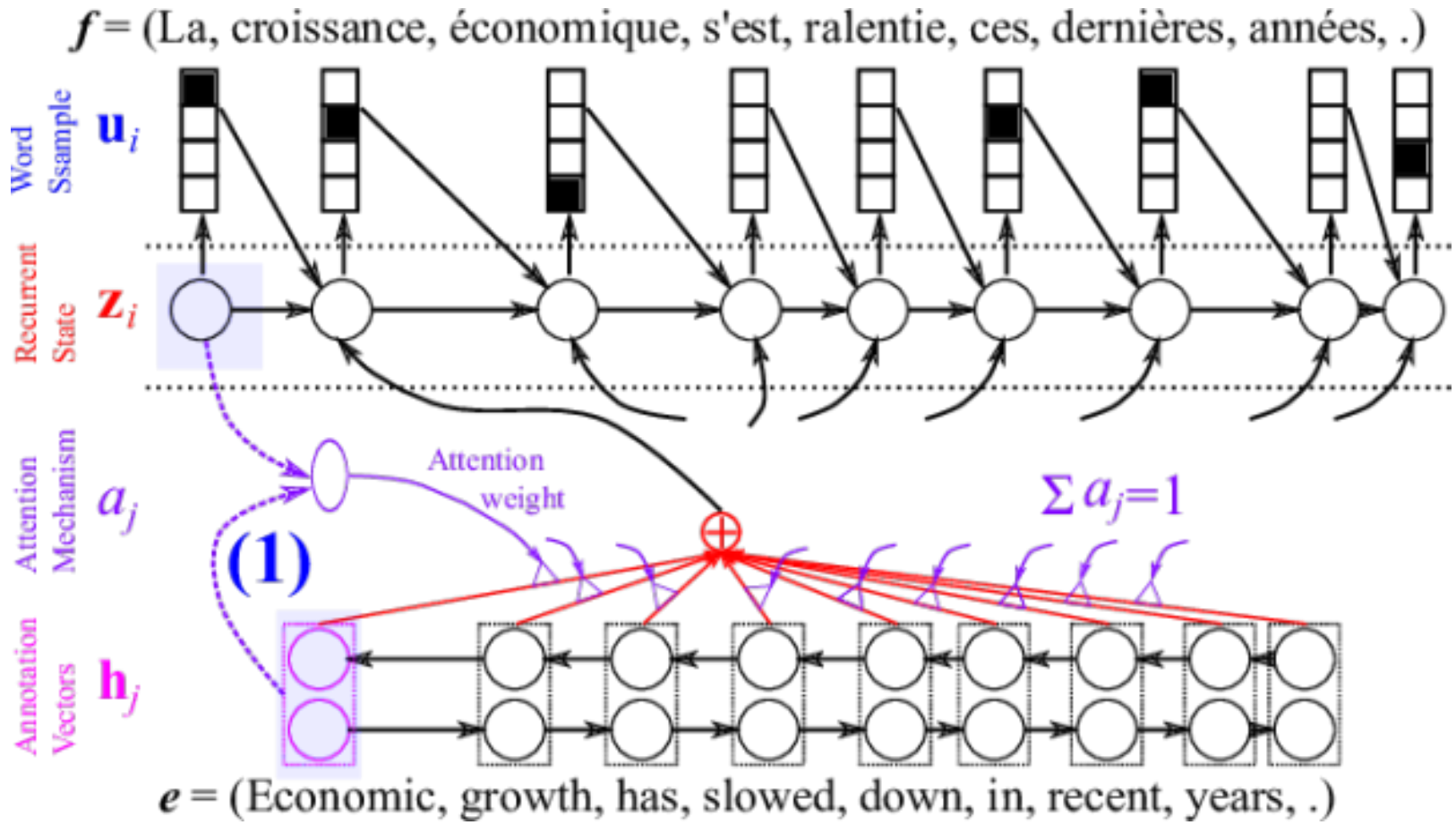
Attention Modeling

Decoding



:: Decoding 시 사용되는 Encoding 정보를 선별적이고 동적으로 바꿔줌으로써 (중요한 것에 집중-Attention함으로써) Decoding 을 더 잘 할 수 있게 됨

Attention Modeling

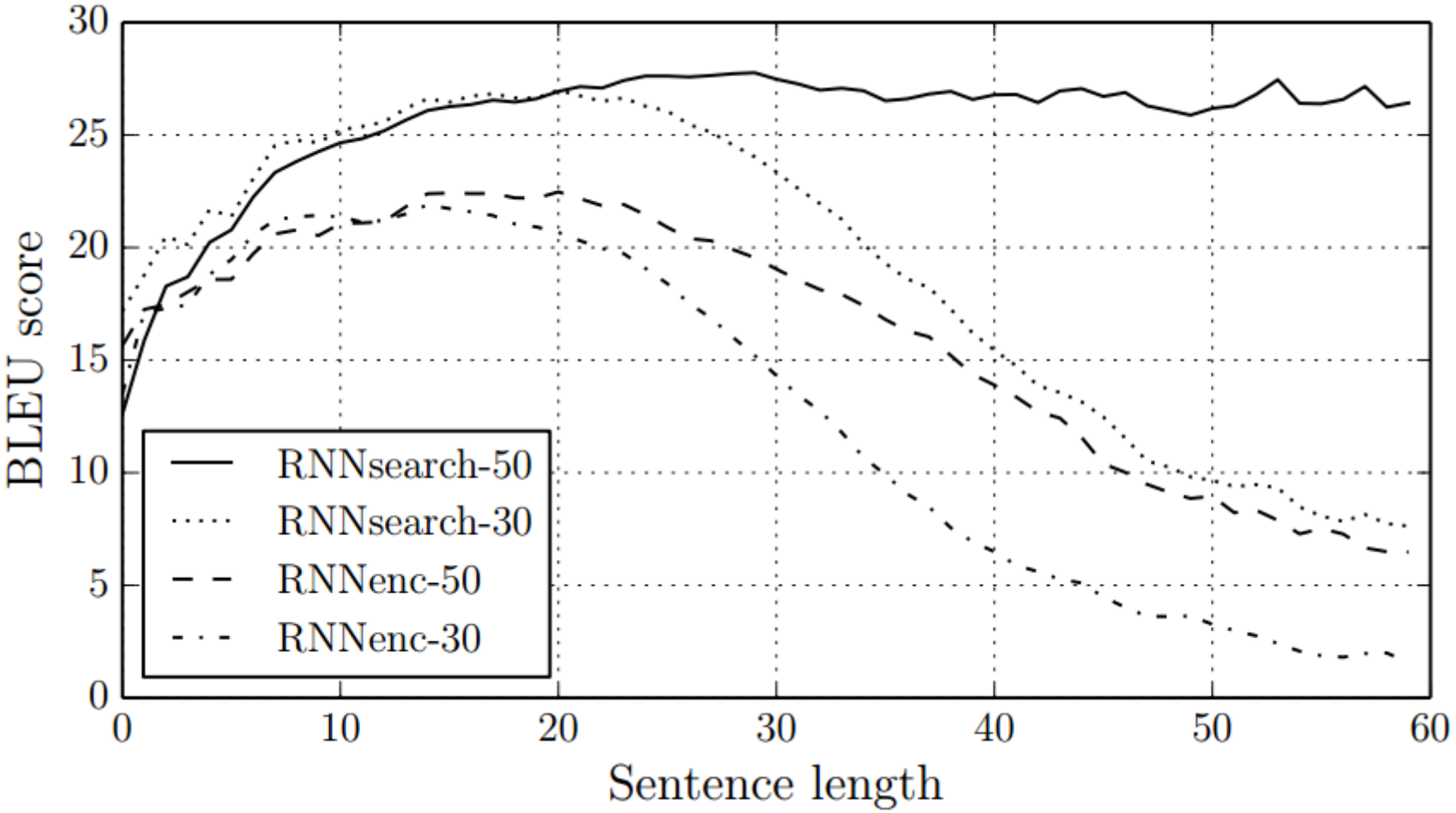


- Bidirectional RNN for Encoding
- Attention Modeling

Bahdanau et al. (2014)

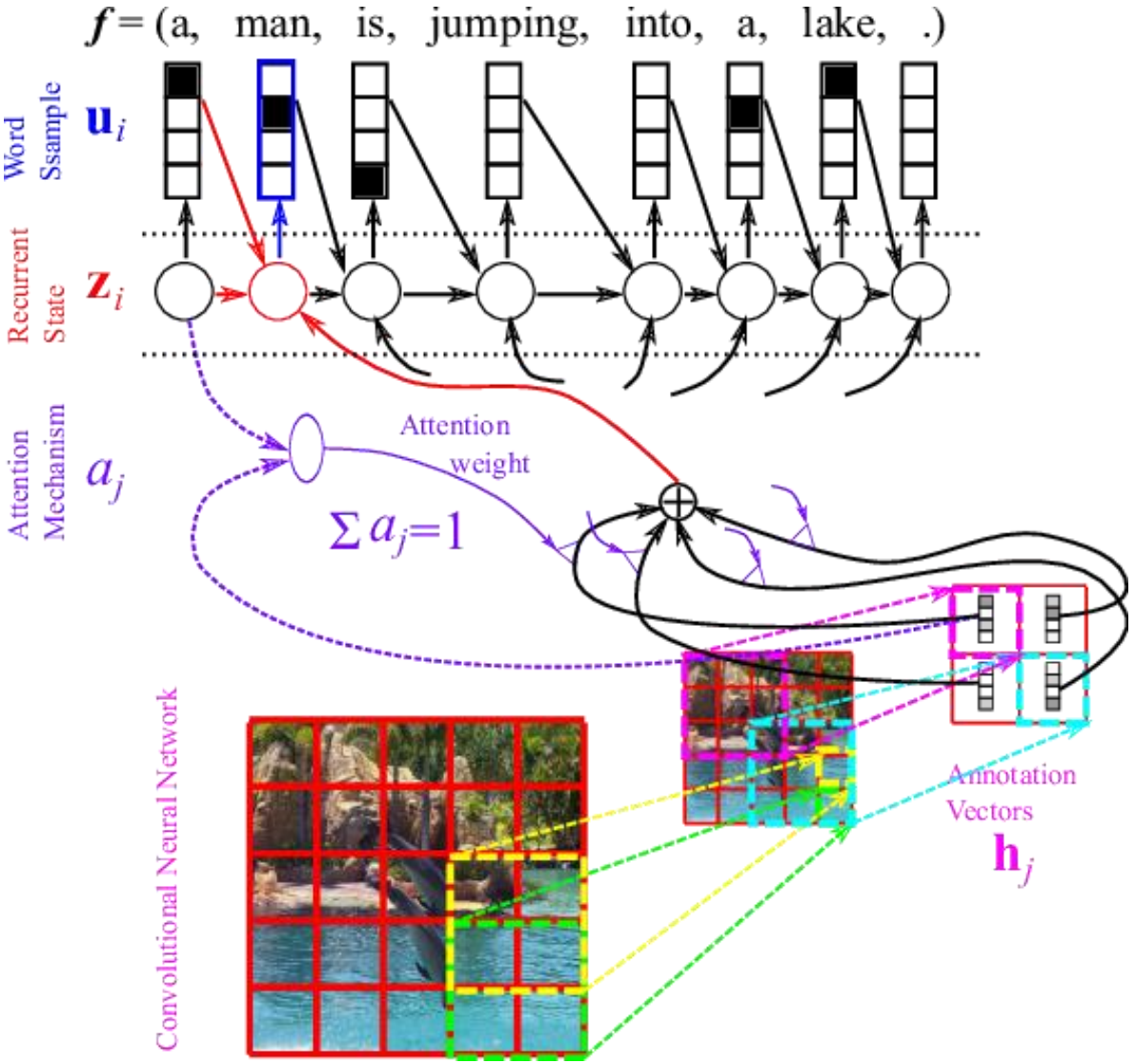
http://devblogs.nvidia.com/parallelforall/introduction-neural-machine-translation-gpus-part-3/

Performance – Attention Modeling @ Machine Translation



:: 선별적으로 가중치가 적용된 Encoding 이 적용됨으로서, 긴 문장에서도 번역 성능이 떨어지지 않는다.

Attention Modeling for Image2Text



http://devblogs.nvidia.com/parallelforall/introduction-neural-machine-translation-gpus-part-3/

Xu et al. (2015)

Show, Attend and Tell: Neural Image Caption Generation with Visual Attention

Attention Modeling for Image2Text

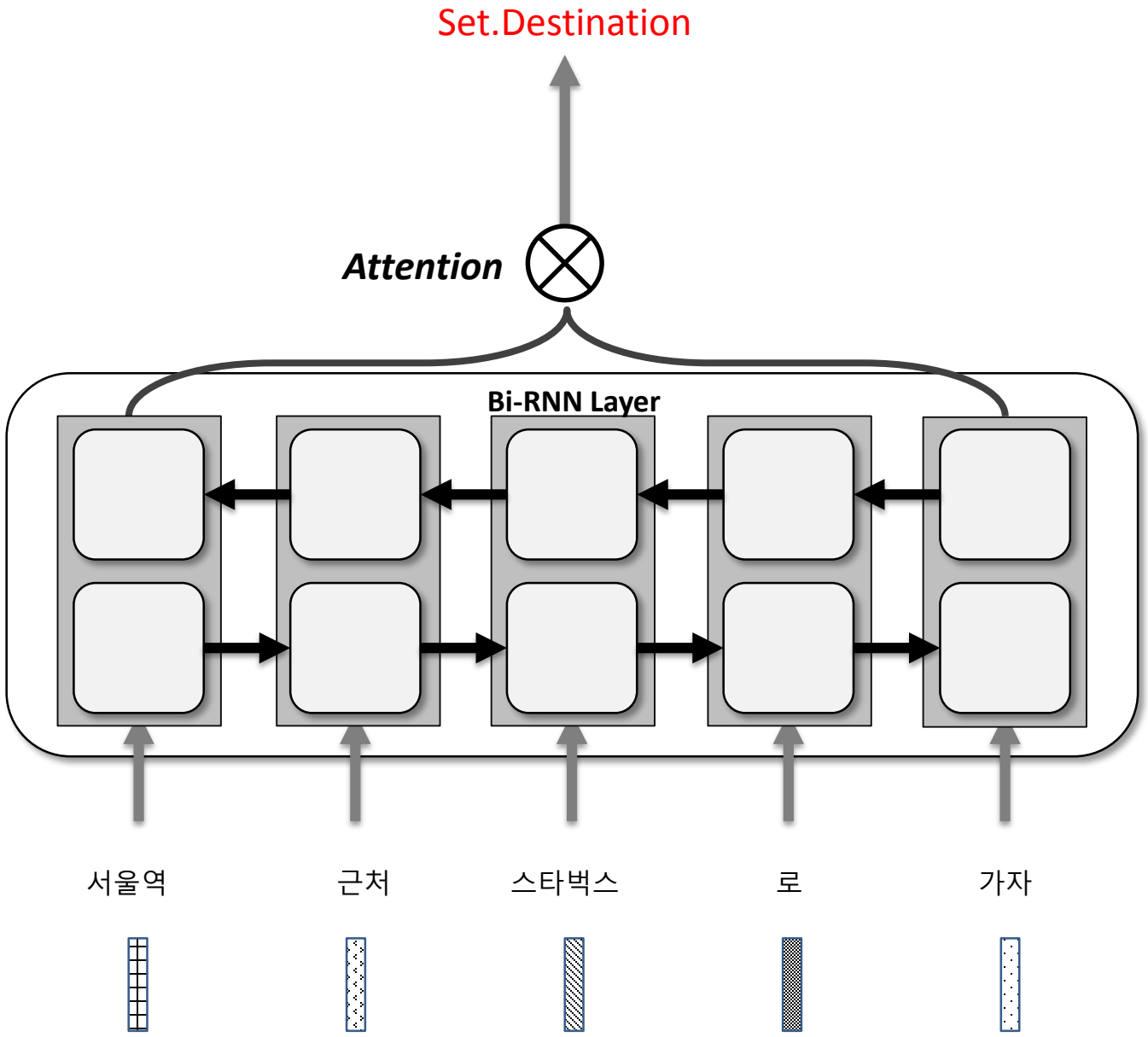
Encoder / Decoder 에서 Text Sequence Encoding 을
Image Sequence Encoding 으로 교체만 해도 똑같이 작동함

A woman is throwing a frisbee in a park.

A little girl sitting on a bed with a teddy bear.

Xu et al. (2015)
Show, Attend and Tell: Neural Image Caption Generation with Visual Attention

Attention Modeling on NLU



정상근

Attention Visualization on NLU

:: 파란색에 가까울 수록 Attention 이 높음

정상근

Demonstration

[자연어이해]

THANK YOU

JOIN THE CONVERSATION

#GTC15

정상근, Ph.D

Intelligence Architect
Senior Researcher, AI Tech. Lab.
SKT Future R&D
Contact : hugmanskj@gmail.com,
hugman@sk.com